

As bibliotecas digitais e o acesso à informação multilingue

Mirijam Sandra Marques Garcia

**Dissertação de Mestrado em Ciências da Informação e da
Documentação - Biblioteconomia**

Outubro 2013

Dissertação apresentada para cumprimento dos requisitos necessários
à obtenção do grau de Mestre em Ciências da Informação e da Documentação
– Biblioteconomia realizado sob:

Orientação Científica:

Doutora Maria de Lurdes Rosa

Co-orientadora:

Mestre Paulo Jorge de Oliveira Leitão

AGRADECIMENTOS

Agradeço em primeiro lugar ao meu orientador o Mestre Paulo Leitão que acompanhou todos os passos da elaboração desta dissertação e sempre mostrou grande disponibilidade e apoio mesmo nas alturas mais difíceis. Agradeço igualmente à minha orientadora a Doutora Maria de Lurdes Rosa pelo incentivo e pela ajuda.

Ainda agradeço a Doutora Rute Costa por ajudar no desenvolvimento desta tese e pelo seu tempo dispensado.

Por fim agradeço a minha família, aos meus amigos e ao meu namorado que incomodei com esta dissertação mais que uma vez ao longo dos últimos meses.

Agradeço ainda à Biblioteca de Arte da Fundação Calouste Gulbenkian me disponibilizou o tempo necessário para realizar este projeto.

As bibliotecas digitais e o acesso a informação multilingue

Resumo

A presente dissertação trata do tema do multilinguismo em bibliotecas digitais, destacando a importância da temática no âmbito do acesso e da RI (Recuperação da Informação). O objetivo é a consideração e elaboração de uma forma de análise das funcionalidades multilingues disponibilizadas, aplicável a estudos de caso. Este estudo é iniciado por uma exposição genérica da temática das bibliotecas digitais e do multilinguismo, com a explicitação de conceitos considerados relevantes e determinantes para este campo de estudo. De seguida, são consideradas algumas problemáticas atuais e soluções relacionadas com o tema: a relação desta área com o campo da terminologia e alguns pontos fulcrais no que toca à avaliação no acesso e na recuperação da informação multilingue. Após esta exposição teórica e a observação exploratória dos estudos de caso selecionados pretende-se, então, criar critérios de análise destinados à avaliação e à comparação de bibliotecas digitais multilingues. Por último, após a aplicação dos critérios, serão pormenorizadamente apresentados e comparados os três estudos de caso selecionados de bibliotecas digitais multilingues apresentando, simultaneamente, os resultados obtidos. Pretende-se desta forma desenvolver um sumário de boas práticas através do estudo efetuado. Estas boas práticas destinam-se à identificação de incoerências, apresentando igualmente áreas de melhoria. Em todo este estudo o utilizador é o ponto fulcral, uma vez que a análise aqui efectuada centra-se no acesso à informação multilingue através da perspectiva do público da biblioteca digital.

PALAVRAS-CHAVE: Bibliotecas digitais, Multilinguismo, Acesso e recuperação da Informação

Digital libraries and multilingual information access

Abstract

The present dissertation addresses the theme of multilingualism in digital libraries, highlighting its importance in the context of information access and retrieval. This study aims, in addition of emphasizing the themes referred, to regard ways of analyzing the multilinguistic features offered applicable to case studies. This review is initiated by a general presentation of the topic of digital libraries and multilingualism, specifying the concepts considered relevant and crucial in this study field. Subsequently are considered some current issues of multilingual information access and retrieval; the relation between this area and the field of terminology, as well as key points of multilingual information access and retrieval assessment. The goal, after this analysis and a previous exploratory observation of the selected case studies, is the development of assessment criteria to evaluate and compare multilingual digital libraries. Finally, after applying the criteria and comparing the three case studies of the digital libraries and their multilinguistic features, the objective is to develop a summary of good practices through the results obtained. These good practices intend to identify inconsistencies and areas that need improvement. Throughout this study the user is a central figure. The assessment aims to focus on multilingual information access and retrieval from the perspective of the digital library public.

KEYWORDS: Digital libraries, Multilingualism, Information Access and Retrieval

Lista de abreviaturas, siglas e acrónimos

CLIR – Cross-language Information Retrieval

ICDL – International Children’s Digital Library

MLIA – Multilingual Information Access

MLIR – Multilingual Information Retrieval

MT – Machine Translation

OCR – Optical Character Recognition

OPAC – Online Public Access Catalog

RI – Recuperação da informação

WDL – World Digital Library

Índice

Introdução	1
1. Bibliotecas digitais e o multilinguismo: uma área de estudo	4
1.1. O multilinguismo em bibliotecas digitais	4
1.2. Recuperação da informação: <i>Cross-language Information Retrieval</i> (CLIR) e <i>Multilingual Information Retrieval</i> (MLIR)	6
1.3. Multilinguismo e o acesso de informação: Aspectos terminológicos	10
1.4. Principais problemáticas no âmbito da RI e do multilinguismo	13
1.5. Critérios de avaliação para bibliotecas digitais multilingues	17
2. Abordagem metodológica	23
2.1. Estudos de caso: critérios de seleção	25
2.2. Casos selecionados: caracterização	27
2.2.1. O caso Europeana	27
2.2.2. O caso World Digital Library (WDL)	29
2.2.3. O caso International Children's Digital Library (ICDL)	30
2.3. Critérios de análise	32
3. Análise dos resultados	39
3.1. Análise e recolha de dados dos estudos de caso	39
3.2. Análise comparativa	50
4. A rentabilização do multilinguismo em bibliotecas digitais	55
Conclusões	61
Referências bibliográficas	63
Anexo I: Interface: ICDL	I
Anexo II: Interface: ICDL	II
Anexo III: Interface: Europeana	III
Anexo IV: Interface: Europeana	IV
Anexo V: Interface: WDL	V
Anexo VI: Interface: WDL	VI
Anexo VII: Pesquisa e Navegação (Pesquisa avançada): ICDL	VII
Anexo VIII: Representação dos resultados: Europeana	VIII
Anexo IX: Representação dos resultados: WDL	IX
Anexo X: Representação do recurso: ICDL	X
Anexo XI: Representação do recurso: ICDL	XI

Anexo XII: Representação do recurso: Europeana	XII
Anexo XIII: Representação do recurso: WDL	XIII
Anexo XIV: Funcionalidades Web 2.0: ICDL	XIV

Introdução

Pretende-se com o presente trabalho abordar a temática do multilinguismo no âmbito das bibliotecas digitais. O ambiente mais frequente desta tipologia de bibliotecas é a internet que conta com inúmeros utilizadores com necessidades de informação diversas, nomeadamente no campo multilinguístico. A importância do estudo prende-se com a necessidade de identificar critérios de análise que permitam explicitar a importância de funcionalidades de acesso e recuperação da informação multilingue.

As bibliotecas digitais podem ser consideradas como a tipologia do futuro, visto que os projetos neste âmbito são numerosos e é possível identificar um crescente interesse na produção de conteúdos digitais e no acesso e recuperação da informação. Destaca-se ainda a crescente investigação neste campo de ação. O aspeto do multilinguismo é essencial uma vez que, atualmente, a informação necessita de ser difundida com as mínimas restrições possíveis tendo como foco principal o utilizador. Devido a esse facto, é determinante o estudo e a análise da pertinência e da eficácia dos mecanismos que levam a que o utilizador de uma biblioteca digital possa usufruir de um serviço completo que corresponda aos padrões criados pela atual Sociedade da Informação. Um ponto fulcral neste âmbito é o estabelecimento de critérios de análise que permitam uma comparação. Desta forma, podem ser identificadas boas práticas relevantes no âmbito dos estudos de caso escolhidos que permitam responder as necessidades do utilizador multilingue e dar resposta às necessidades de informação deste.

Em termos de relevância considera-se que é importante realizar este estudo pois trata-se de uma área crescente devido ao importante papel do tratamento e da disponibilização da informação no meio digital, ou seja, a internet. Essa informação não se restringe à produção inglesa, a presença de outras línguas é cada vez mais frequente e determinante. As bibliotecas digitais são uma tipologia em crescimento constante com uma grande produção de conteúdos disponibilizados e novos projetos. Numa Sociedade de Informação global, o acesso à informação multilingue é uma obrigatoriedade para combater a exclusão de informação e para criar igualdade nas possibilidades de acesso e recuperação da informação.

No que respeita à significância do estudo, ou seja, como este estudo contribui para a comunidade científica, esta exprime-se no salientar de boas práticas através da criação de um quadro de avaliação e comparação que se centra no acesso e na

recuperação da informação multilingue. Através da criação deste quadro pretende-se destacar a importância e o pormenor de uma análise na perspetiva do utilizador, de forma a proporcionar uma experiência flexível no espaço digital e no acesso aos vários conteúdos disponibilizados.

O objetivo central deste estudo é analisar as funcionalidades multilingues numa biblioteca digital de carácter global multilingue, ou seja, a forma como os utilizadores acedem e recuperam a informação num espaço digital que disponibiliza várias línguas e quais as possíveis problemáticas com as quais se deparam. Para poder analisar essas funcionalidades, ou seja, as formas de acesso e recuperação de informação, foram selecionados critérios através da revisão de literatura e de uma fase prévia de observação. Os critérios resultantes funcionam como sugestão de análise da recuperação e do acesso à informação em ambientes específicos multilingues. Nesta análise também se pretende integrar a existência das funcionalidades da Web 2.0. Estas funcionalidades podem contribuir para a criação e melhoria da informação no âmbito do acesso e da recuperação da informação.

Um dos pontos fulcrais, passa ainda por, salientar o carácter interdisciplinar da ciência da informação e demonstrar a relação que existe entre o campo da ciência da informação, tendo em conta, especialmente, o acesso e a recuperação da informação com a área da terminologia. A terminologia é um fator fundamental que permite assegurar a qualidade, eficácia e eficiência para qualquer sistema de RI multilingue. Contribui também para o seu funcionamento adequado e isso reflete-se nos benefícios que pode trazer para os utilizadores destes sistemas.

No primeiro capítulo pretende-se realizar uma exposição teórica dos principais temas, podendo destacar a aplicação do multilinguismo às bibliotecas digitais e a sua relevância. Trata-se, igualmente, neste capítulo das funcionalidades e dos conceitos relacionados com o acesso e a recuperação multilingue, englobando problemáticas e possíveis soluções. É ainda abordada a influência que os aspetos terminológicos podem ter em termos de qualidade nestes sistemas. No capítulo seguinte, e baseada na exposição teórica anterior, desenvolveu-se uma abordagem metodológica que permitiu a criação de critérios de análise. Os critérios foram obtidos a partir de uma junção de aspetos teóricos e de uma observação exploratória dos estudos de caso selecionados. Os critérios de análise foram divididos pelas fases do processo da recuperação da

informação e agrupados em grelhas. Estes critérios, a sua aplicação e os resultados foram salientados no capítulo 3, procedendo-se a uma análise faseada, tratando-se de cada caso individualmente e respeitando a ordem das grelhas de análise. No ponto seguinte efetuou-se uma análise comparativa entre os três casos de estudo selecionados. A partir desta análise foi possível identificar boas práticas. No último capítulo pretende-se uma junção de todos os pontos destacando os pilares principais nos quais o estudo foi desenvolvido e questões que são importantes para a rentabilização do multilinguismo em bibliotecas digitais.

1. Bibliotecas digitais e o multilinguismo: uma área de estudo

1.1. O multilinguismo em bibliotecas digitais

No contexto europeu pretende-se uma sociedade inclusiva na qual seja inerente a compreensão mútua entre diferentes culturas. O objetivo é construir uma identidade europeia comum a todas as suas partes integrantes. Mais especificamente no domínio da língua, esta reflete a identidade de um país, faz parte da sua cultura e é parte integrante de todos os sistemas de informação que se desenvolvem nacional e internacionalmente. A língua não pode ser ignorada quando se trata de aspetos culturais. (MINERVA, 2006, p.13) O âmbito europeu de multiculturalismo e de multilinguismo representa uma grande oportunidade que as instituições culturais devem explorar para se poderem colocar na esfera digital da sociedade da informação. (MINERVA, 2006, p. 15) As bibliotecas, em geral, podem ser consideradas espaços multilingues. Normalmente trata-se de um ambiente, preenchido de documentos nas mais variadas línguas, no qual é oferecido o acesso aos mesmos, tal noção é ainda mais crucial para as bibliotecas digitais. Essa importância destaca-se nas formas de acesso ao documentos disponibilizados que necessitam ter igualmente características multilingues para servirem como auxílio aos utilizadores. (MAEDA, 2002, p.1) Bibliotecas digitais multilingues são bibliotecas digitais que providenciam o acesso a conteúdos em mais do que uma língua ou então o acesso multilingue através de *queries* a uma coleção monolingue. O objetivo essencial prende-se com a capacidade de acesso à informação a uma escala global. (DIEKEMA, 2012, p.1)

Uma biblioteca digital no contexto do multilinguismo é caracterizada por ter todas as funções de acesso à informação implementadas de forma simultânea em tantas línguas quanto as desejadas. Outros fatores que definem esta tipologia são as funções de pesquisa e de recuperação da informação independentes da língua. A implementação de tal conceito permite ao utilizador escolher a língua para o interface, independentemente do idioma utilizado no contexto documental na biblioteca. (PAVANI, 2001, p.73) Desta forma, para responder às questões da crescente globalização e do crescimento do mundo da informação na internet, as bibliotecas digitais têm sido criadas não apenas para superar as barreiras físicas, mas também para superar as barreiras de natureza linguística. (DIEKEMA, 2012, p.1) No limite, uma biblioteca digital global teria a capacidade de possibilitar o acesso a informação em todas as línguas existentes. Neste

seguimento, os utilizadores poderiam aceder à informação não só na sua língua nativa mas também a informação traduzida nas línguas pretendidas pelo próprio, em vários formatos. Apesar deste tipo de bibliotecas ainda não constituir uma realidade, a investigação destas formas de recuperação da informação é determinante nomeadamente para o desenvolvimento de futuros modelos de plataformas digitais. Uma biblioteca digital global apresenta várias possibilidades, como por exemplo tornar os recursos de informação mais acessíveis às várias comunidades de utilizadores, mesmo as ainda desconhecidas, permitindo ainda o acesso pelos utilizadores a vários formatos, línguas e representações. (BISWAS, 2005, p.176)

O aumento do número de recursos que apenas se encontram disponíveis na língua inglesa na internet é nítido. Desta forma é necessário desenvolver técnicas de recuperação de informação para as outras línguas. Este desenvolvimento é um desafio recente e de natureza urgente. No âmbito da internet podem ser referidos vários motores de pesquisa que já oferecem o seu serviço em várias línguas, sendo a Google o exemplo paradigmático, uma vez que este sistema suporta uma grande variedade de línguas. Também as funcionalidades de tradução automática estão aqui presentes. Existem outras funcionalidades que também permitem a recuperação de informação, assim como a pesquisa no site, nas hiperligações e a tradução da página Web entre outras. No entanto, apesar da existência das mesmas, continua a ser mais eficiente a recuperação de informação no âmbito da língua inglesa. (ZHOU, 2006, p.15-16) Em concordância com o que foi anteriormente referido WOLDERING (2006, 2-3) afirma que os conteúdos na Web não se caracterizam por serem monolíngues, também o número de conteúdos em outras línguas se encontra em rápido crescimento. O perfil dos utilizadores que já não se centra apenas no âmbito académico tem igualmente vindo a tornar-se mais abrangente. Estes fatores levam a uma pressão crescente no sentido de providenciar o acesso a informação sem restrições linguísticas ou culturais. As bibliotecas, as suas aquisições e os seus públicos nunca foram monolíngues, algo que se tornou ainda mais evidente desde que as bibliotecas começaram a oferecer os seus serviços no espaço Web. (WOLDERING, 2006, pág. 2-3) Assim, as Bibliotecas digitais poderão responder de melhor forma aos seus utilizadores se forem implementados serviços de suporte multilinguístico nos seus sistemas. (CHEN; BAO, 2009, p.8)

Um elemento determinante em todos os sistemas de informação, incluindo bibliotecas digitais, é sem dúvida a pesquisa. É necessário considerar que grande parte

da informação armazenada em bibliotecas digitais não é acessível para motores de pesquisa e, como tal, trata-se de um benefício geral o facto de as bibliotecas disponibilizarem serviços de acesso multilinguísticos. Para tal CHAO e BAO (2009, p. 7-8) destacam quatro formas de implementação para esse efeito. Em primeiro lugar a colaboração com investigadores na área, colaboração com criadores de bibliotecas digitais estrangeiras e com utilizadores, assim como uma abordagem passo a passo: primeiro desenvolvendo o interface, depois desenvolvendo os metadados e, por fim, desenvolvendo a coleção inteira, permitindo o acesso multilinguístico.

Para melhor abordar o princípio da pesquisa é necessário destacar as formas como a recuperação e acesso da informação podem ser feitas em bibliotecas digitais, quais os sistemas mais frequentes e o que é que esta questão implica.

1.2. Recuperação da informação: *Crosss-language Information Retrieval (CLIR)* e *Multilingual Information Retrieval (MLIR)*

O META TECHNOLOGY COUNCIL (2012, p.7) afirma que menos de 10% da população europeia tem a capacidade de usufruir dos serviços da internet em inglês. Esta razão torna mais essencial a disponibilização de serviços e funcionalidades multilingues. O objetivo é a criação de um mercado digital europeu comum que conta com mais de 500 milhões utilizadores.

A internet revolucionou sem dúvida a RI, cada vez mais são utilizados sistemas de recuperação mais sofisticados como o Google ou sistemas criados para o uso específico no âmbito das bibliotecas. (SUJATHA; DHAVACHEVAN, 2011, p.115) Permitir o acesso e a recuperação da informação em várias línguas contribui para a construção de uma infraestrutura informacional globalizada. Os problemas afetam tanto o comércio eletrónico, como instituições informacionais, tais como museus, arquivos, bibliotecas, instituições académicas e ainda instituições de hardware, software e aplicações da área de redes. (BISWAS, 2005, p.177)

O papel de RI é a entrega de informação relevante tendo em conta as necessidades demonstradas pelo utilizador em determinado momento. A confrontação com RI é diária, especialmente no âmbito da internet. Esta expansão estimula o desenvolvimento de métodos multilinguísticos para RI, sendo que este desenvolvimento é necessário para ultrapassar as barreiras linguísticas que existem neste campo do acesso

à informação. (SORG, 2011, p.1) GÖDERT (2007, p.1) estabelece um número de funcionalidades desejadas para um sistema de RI. O *matching* de palavras presentes em Web OPAC's e motores de pesquisa - correspondência de palavras - define-se por combinar as palavras utilizadas na *query* com as existentes no documento. No entanto este esquema estandardizado não é considerado suficiente para uma RI eficiente. Uma vez que não evita a ausência de resultados, não permite uma pesquisa temática e nem uma interação intuitiva por parte do utilizador tendo em conta os interesses deste. Ainda considerado determinante é a pesquisa por conceitos que procura alargar a pesquisa por palavra à sua temática. Esta opção permite reduzir, por exemplo, as pesquisas sem resultados. Para este efeito são estabelecidas relações de sinonímia de ficheiros de autoridade que são incluídos no processo de pesquisa. O objetivo será a eficiência dos resultados da pesquisa, de modo a que tenham em conta o âmbito semântico do conceito pretendido e que tal âmbito semântico seja manipulável pelo utilizador. Regra geral, num sistema de RI, o utilizador expressa a sua necessidade de informação através de uma *query*, seguidamente o sistema de RI recupera itens “relevantes” para satisfazer a necessidade de informação do utilizador. A noção de relevância é essencial para a RI, pois a relevância é definida a um nível semântico, desta forma explicita se o conteúdo do recurso recuperado satisfaz a necessidade de informação do utilizador. Outro conceito importante, intrinsecamente relacionado com a rapidez e relevância dos itens recuperados, é a satisfação do utilizador. A definição dada pelo autor é: considerando a coleção “D” contendo recursos de informação “d” e introduzindo uma *query* “q”, RI é definido como a tarefa de recuperação de uma lista hierarquizada dos recursos do tipo d. Sendo que estes estão organizados consoante as necessidades de informação especificados em “q”. (SORG, 2011, p.8/9)

SORG (2011, p.9) define CLIR como a ação de recuperar documentos relevantes a partir de uma determinada *query* numa determinada língua de um conjunto de documentos noutra língua. Por outro lado MLIR considera *corpora* que englobam documentos em diferentes línguas. Neste âmbito a tradução assistida por máquina pode ainda ajudar a traduzir o texto de partida para a língua nativa do utilizador. *Cross-language Information Retrieval* (CLIR) é considerada uma técnica de RI que garante a recuperação de documentos numa determinada língua, usando uma língua diferente na fase da pesquisa, ou seja, na *query*. No entanto, é difícil atingir uma RI eficiente quando se trata de grandes coleções contendo diversas línguas e domínios pois tornam todo o

processo mais complexo, necessitando de formas para recuperar esses recursos. Trata-se de uma técnica de RI utilizada em bibliotecas digitais e na internet em si, sendo necessário melhorar o seu desempenho, o que é uma tarefa complexa no universo no qual se insere. (MAEDA, 2002, p.1) SUJATHA e DHAVACHEVAN (2011, p.115) definem CLIR como uma forma da recuperação de informação que permite formular uma questão numa língua e recuperar documentos em várias línguas. Enquanto MLIR é definido como a colocação de uma questão numa ou mais línguas e a recuperação de documentos numa ou mais línguas. Segundo ZHOU (2006, p.17) a abordagem *Multilingual information retrieval* (MLIR) que é uma extensão de *Bilingual Information Retrieval*, apresenta-se de forma transversal como *Cross-lingual Information Retrieval* (CLIR). ARCHANA e DEVI (2011, p.29-30) afirmam que de CLIR existem três abordagens principais, entre elas a tradução assistida por máquina, vocabulários controlados e a abordagem baseada em dicionários.

A primeira abordagem, a tradução assistida por máquina, pode ser implementada de duas formas: a primeira traduzindo os próprios documentos para a língua da *query* introduzida pelo utilizador e a segunda traduzindo a *query* para efeitos de pesquisa. Neste último caso, levanta-se com a problemática dos erros de tradução devido à falta de informação no índice de termos ou devido a definições ambíguas de conceitos. Muitas vezes também não cobre todos os domínios com a mesma qualidade de tradução, sendo mais eficiente quando envolve terminologia técnica específica. A segunda abordagem que refere o vocabulário controlado é considerada pelo autor como a mais tradicional, esta é usada na sua maioria para efeitos de indexação e recuperação. Neste ambiente são seleccionados pelo documentalista alguns descritores de uma lista controlada de termos. Desta forma é criado um tesouro multilinguístico com os descritores e as relações semânticas. Os descritores podem ser inseridos no tesouro de forma automática ou manual. A problemática aqui representada é que a *query* necessita de usar os termos presentes no tesouro. A terceira abordagem baseada em dicionários é a leitura de cada termo presente na *query* do utilizador sendo controlada pelo sistema num dicionário bilingue legível por máquina. Um dicionário bilingue legível por máquina pode ser definido como uma estrutura de dados que contém uma lista de entrada para um dado conjunto de termos e um mecanismo de verificação, que tendo em conta a língua de partida do termo da *query*, consulta essa estrutura de dados obtendo uma ou mais traduções possíveis ou mesmo equivalentes do próprio termo. A tradução

encontrada é então adicionada ao mapeamento semântico da língua do documento. Esta abordagem necessita menos tempo de implementação.

Dentro destas várias abordagens existem vários métodos de recuperação de informação cujo funcionamento é dependente da abordagem segundo a qual funciona o sistema, cada método de RI possui métodos de tradução. Multilingual Information Access (MLIA) refere-se a tecnologias que permitem ao utilizadores de um sistema recuperar e utilizar informação de uma coleção de documentos multilingues. O foco principal recai, atualmente, nas línguas não muito comuns na qual a informação está apresentada, pois o seu tratamento é mais complexo. A investigação nesta área começou por aproveitar as capacidades de *Cross-Language Information Retrieval* (CLIR). MLIA envolve sempre CLIR, pois permite a recuperação de documentos numa língua diferente da qual a *query* está formulada. Nesta técnica podem ser destacadas então os três métodos de tradução. A primeira, a tradução da query (que traduz a query do utilizador para a língua na qual se encontram os documentos pretendidos), tradução do documento (traduz a coleção dos documentos disponíveis para língua do utilizador) e a abordagem interlíngua que tira proveito das duas últimas abordagens referidas. (Chen [et al.], 2012, p.1,2) As três ferramentas mais utilizadas para esse efeito são dicionários, *corpora* paralelos e sistemas de tradução assistida por máquina. Quando se trata de *queries* as ferramentas mais utilizadas são as primeiras duas, já para os documentos a mais utilizada é a tradução assistida por máquina. No caso da tradução baseada em dicionários trata-se de uma lista de palavras que contém tanto a língua de partida como a língua de chegada. A tradução assistida por máquina é utilizada para traduzir a *query* ou os documentos. As maiores dificuldades identificadas prendem-se com os gastos em termos de implementação do sistema e o facto de também não ser aplicável a situações com grandes quantidades de documentos. No caso dos *corpora* paralelos estes também se apresentam como bastante dispendiosos, apesar de melhorarem bastante o desempenho da recuperação de informação. No entanto, para línguas menos comuns pode tornar-se muito difícil encontrar *corpora* paralelos. (SUJATHA; DHAVACHEVAN, 2011, p.115-116) A tradução é uma tarefa desafiante pois não é possível mapear duas línguas diferentes conjuntamente e de forma perfeita. Uma determinada palavra numa língua pode ter vários significados noutra, sendo que este se encontra intrinsecamente ligado ao seu contexto e por essa razão a tradução automática é tão limitada. Na recuperação monolíngua algumas dificuldades são amenizadas por

existirem formas estandardizadas. Podem citar-se os exemplos de formas plurais e singulares e as raízes das palavras, vocabulários controlados e ainda um tesauro pode ser utilizado com o intuito de estabelecer formas preferenciais e referenciar a utilização de sinónimos ou termos relacionados. Ainda podem ser utilizados algoritmos que tenham informação sobre a estrutura gramatical e a frequência de palavras. (BISWAS, 2005, p.177)

1.3. Multilinguismo e o acesso de informação: Aspetos terminológicos

Um dos pontos fulcrais é o destaque de como o carácter interdisciplinar da Ciência da Informação pode beneficiar em termos qualitativos o acesso e a recuperação da informação. Pretende-se explicitar a interligação entre os aspetos da biblioteconomia, nomeadamente no que diz respeito à organização, ao acesso e recuperação da informação, com a terminologia. A terminologia é uma área interdisciplinar e tem como objeto primário a análise de conceitos e de sistemas conceptuais assim como a criação, deteção e extração de termos novos ou já existentes. Outro campo de ação da terminologia é a compilação e seleção de recursos terminológicos. (GORNOSTAY; [et al.], 2012, p.1) As ferramentas multilingues dependem de funcionalidades de tradução que, por sua vez, se encontram intrinsecamente ligadas à terminologia. A terminologia é essencial para o desenvolvimento de sistemas de acesso e de recuperação de informação. A maioria das bibliotecas digitais no âmbito internacional necessitam como já referido, de tesauros, sendo neste caso especificamente um tesauro multilingue que organiza o conhecimento através de mapas conceptuais para assim garantir o acesso e a recuperação da informação. Com a ajuda do tesauro é possível estabelecer uma normalização ao organizar o vocabulário nas várias línguas e desta forma estabelecer ligação. (MENDES, 2008, p.1; p.7) No entanto, é necessário adaptar o tesauro multilinguístico às necessidades de uma biblioteca digital, tendo em conta as línguas que esta necessita para a recuperação de informação e as especificidades de cada língua. Também as ontologias como SILVA (2012, p.15-p.18) afirma, apesar de consideradas independentes das línguas utilizadas, tanto o processo de criação como o de acesso a informação é feito com o auxílio a língua, ou seja, trata-se do âmbito do multilinguismo. As ontologias são determinantes no que diz respeito à reutilização da informação e a aspetos de interoperabilidade. SILVA (2012, p.18-19) identifica que é necessária a criação e o aperfeiçoamento de recursos multilingues, os quais são identificados como “recursos previamente estruturados”, tais como “classificações, tesauros, *wordnets*,

taxonomias ou ontologias”. No caso destes recursos, a informação de um determinado domínio encontra-se já de forma organizada e está estruturada consoante as especificidades de necessidade de cada tipologia de recurso, podendo todas ser adaptadas às particularidades do multilinguismo. É ainda necessário destacar a proximidade que existe entre sistemas terminológicos e sistemas de classificação o que se reflete, principalmente, nas funcionalidades de pesquisa e na organização da informação. Também em termos de funcionalidades multilingues estas podem ser consideradas baseadas em sistemas terminológicos. Sistemas de classificação acompanhados de terminologia fidedigna e terminologia multilingue para permitir funcionalidades multilingues permitem a organização dos conteúdos, a criação de novos sistemas e a otimização e gestão do processo documental, incluindo aspetos de tradução. Isto torna um sistema mais eficiente. (SAUBERER, 2011, p.3) Referir-se à organização da informação num contexto cada vez mais global torna aspetos terminológicos cada vez mais prementes. Como SILVA (2012, p.20) afirma “A existência de terminologias de carácter multilingue, reutilizáveis e partilháveis, é, assim, indispensável à existência de uma comunicação especializada rigorosa e eficaz no seio da sociedade de informação, dado o papel que representam enquanto meios de acesso e disseminação do conhecimento.”

É cada vez mais necessário considerar o papel do utilizador. Cada biblioteca tem um determinado público-alvo que é de extrema importância, tanto como elemento de interação que dá uso à informação, como noutra perspetiva que permite visualizá-lo como cocriador da própria informação, transferindo-o de um papel passivo, recetor da informação, para criador de informação. O público-alvo de uma biblioteca digital não estará, normalmente, familiarizado com a terminologia específica de uma determinada área. Assim, conceitos específicos da área da ciência da informação, como por exemplo os próprios “metadados” podem ser desconhecidos. Levanta-se então a questão se tal palavra deve constar numa biblioteca digital ou se a designação dos campos dos metadados escolhidos para designar determinada realidade descritiva é absorvida pelo utilizador da forma pretendida pelo gestor de informação que a criou. Como referido por MENDES (2008, p.5) o profissional de informação que gere os conteúdos é responsável pela mediação no que diz respeito ao utilizador e ao acesso e recuperação da informação. A indexação é uma das formas de conseguir estabelecer essa ponte, ao criar conceitos de pesquisa para a gestão da informação presente na biblioteca. SILVA (2012,

p.22) ainda complementa esta ideia referindo que aspetos como a “identificação, aquisição, organização, criação, uso e disseminação” da informação são o campo de ação central da gestão da informação. É necessário o desenvolvimento de processos estratégicos em conjunto com fatores tecnológicos. Este desenvolvimento permite a existência de uma estrutura de apoio a essa mesma gestão, pois é necessária a compreensão de como indivíduos, grupos ou mesmo organizações fazem uso, denominam e disseminam a informação e a sua área de conhecimento. Em relação aos utilizadores de uma determinada biblioteca é ainda importante o que SILVA (2012, p.18-p.19) designa por novas abordagens de caráter multidisciplinar, refere-se as ajudas que as funcionalidades da Web 2.0, nomeadamente das redes sociais podem proporcionar. No entanto, este autor destaca ainda que as redes sociais e as redes colaborativas também geram a necessidade de novas formas de sistemas de organização de informação que consigam aproveitar de forma eficiente a informação que circula nessas redes. Desta forma, iria simplificar-se o acesso e da RI multilingue, por contribuir com um funcionamento mais eficiente e interativo, alargando as possibilidades de reutilização da informação e de aperfeiçoamento da informação já existente.

No que diz respeito aos aspetos de tradução, podemos destacar dois essenciais, por um lado a tradução já referida dos metadados, que tem por alvo a tradução da denominação de campos específicos determinantes para a representação de um recurso digital e, por outro lado, a tradução do texto. Como texto podem ser identificados todos os elementos textuais com informação, presentes em qualquer interface de uma biblioteca digital, para além da designação dada dos campos dos metadados correspondentes a determinado recurso. Qualquer tipologia de biblioteca, excluindo as bibliotecas especializadas, tem a responsabilidade de representar várias áreas do conhecimento, sendo que cada uma dessas áreas possui a sua terminologia de especialidade. Os elementos textuais podem assim representar conceitos específicos dos mais variados domínios. Os metadados representam a linguagem de descrição conceptual específica desenvolvida para representar um objeto digital. SILVA (2012, p.25-p.31) explicita essa importância da identificação de conceitos para ser possível representá-los num sistema de informação. Sem a representatividade do conceito a estruturação da informação é impraticável. É necessário identificar conceitos que integram um domínio específico. Quando lidamos com tradução não é uma simples

tradução do texto de partida para o texto de chegada. É necessário, tendo vários utilizadores, definir bem os conceitos utilizados neste campo de ação para que sejam acessíveis ao público universal. Nunca foi tão importante preocupar-se com terminologia e funcionalidades de tradução como hoje pois influencia toda a eficácia e eficiência da recuperação da informação. Como MENDES (2008, p.58) afirma a terminologia desempenha um papel fulcral tanto na organização como na recuperação da informação e é assim uma parte determinante para a Ciência da informação. SILVA (2012, p.56) afirma que o objeto de estudo da terminologia são “os conceitos e as suas designações, os termos, bem como as relações que aqueles estabelecem entre si, no seio de um domínio de especialidade, estando inextricavelmente ligado ao conceito de conhecimento especializado, à sua organização e à sua representação.” SAUBERER (2011, p.2) sublinha ainda que atualmente na sociedade de informação com o crescimento documental e com o multilinguismo destacar aspetos terminológicos pode evitar fatores de distorções na comunicação e ambiguidades terminológicas. A comunicação entre o desenvolvimento, produção, marketing, organização, parceiros, fornecedores, o mercado e o cliente torna-se mais transparente. É essencial quando se trabalha com conceitos considerar que podem ser encaradas duas dimensões, uma linguística e uma conceptual necessária para compreender e identificar a dimensão de um determinado elemento de comunicação. A representação conceptual é transversal a qualquer língua utilizada. Quando se trata da escolha de um termo representativo este não se deve centrar na simples tradução ou equivalência mas no que é considerado o termo natural utilizado no mesmo contexto no âmbito da outra língua. (SILVA, 2012, p.69-71)

As bibliotecas digitais podem assim considerar utilizar todo um mundo de ferramentas e de métodos criados no âmbito da linguística para melhorar os seus sistemas de acesso e recuperação da informação.

1.4. Principais problemáticas no âmbito da RI e do multilinguismo

No caso das bibliotecas digitais multilingues um dos principais problemas é a barreira linguística. DIEKEMA (2012, p.6-7) afirma que a recuperação da informação *cross-language*, ou seja, entre línguas, necessita de funcionalidades de tradução ou de superar a barreira linguística. Neste caso a maior problemática são os erros que ocorrem através do processo de tradução, o que resulta num impacto negativo em termos do

resultado das pesquisas em bibliotecas digitais. Situações como ambiguidade lexical e sinonímia causam problemas de recuperação de informação mesmo no caso de bibliotecas monolíngues. Existem vários fatores que causam erros na tradução, assim como, a falta de tradução para termos técnicos, acrónimos ou nomes próprios e ainda a adição de múltiplos sensores de tradução de uma palavra. Outro problema que deve ser destacado é o da gestão de dados, a gestão de repositórios e o armazenamento de conteúdos e metadados que possuem aspetos que tornam difícil o trabalho eficaz das bibliotecas digitais. Os vocabulários utilizados nos metadados necessitam de tradução correta, também o interface tem que ser traduzido, e a indexação em várias línguas também se torna desafiante pelas diferentes regras presentes consoante a língua tratada. (BIA et al. cit. por DIEKEMA, 2012, p.7). Mesmo o *Optical Character Recognition* (OCR), que permite o reconhecimento de caracteres num ficheiro de imagem, pode tornar-se um problema para algumas línguas. O interface de uma biblioteca digital multilíngue necessita de funcionalidades que permitam o acesso a todas as línguas oferecidas pelo sistema para permitir o acesso a todos os documentos disponíveis. No entanto também deve incluir outras funcionalidades que permitam a pesquisa e a recuperação multilíngue e *cross-language*. No entanto pode ser problemático encontrar métodos eficazes que combinem as *queries* com documentos transversais a todas as línguas. (BISWAS, 2005, p.179)

São identificados três abordagens de tradução no que se refere a CLIR, a primeira tratando-se de tradução de *queries*, que é considerado o mais utilizado, tradução de documentos e técnicas interlinguísticas. As bibliotecas digitais podem escolher a utilização de Tradução por máquina (MT) para todos os documentos antes de indexar, ao invés de proceder à tradução de *queries* na altura da pesquisa. No caso das bibliotecas digitais com coleções estáveis podem adquirir mecanismos assistidos por computadores para construir a sua base de tradução para a tradução de *queries*. BUDZISE-WEAVER, CHEN e MITCHELL (2011, p. 9) concordam com essa visão identificando CLIR e MT como os principais desafios e oportunidades para bibliotecas digitais pois estas abordagens atualmente disponibilizam de uma tecnologia avançada. Segundo ARCHANA e DEVI (2011, p.28- 30) CLIR envolve três problemas principais: o primeiro é superar a barreira da língua, o segundo envolve determinar qual é o método de tradução mais indicado, sendo mais benéfico para a tradução selecionar mais que um método. O terceiro trata de dar relevância a termos corretos quando mais do que uma

tradução é escolhida. Uma forma de apoiar o processo de resolução destas problemáticas é o *Data mining*. *Data mining* é definido como uma técnica que permite a pesquisa em grandes bases de dados para encontrar padrões. Esta técnica pode automatizar este processo e gerar nova informação e valorizar os recursos de informação já existentes. *Data mining* e RI são tecnologias semelhantes na medida em que as duas permitem a análise, a organização e a pesquisa em documentos textuais, multimédia e informação estruturada e semiestruturada. A criação de informação por meio de *data mining* é considerada um processo de descoberta automática de informação relevante num grande conjunto de dados. Esta técnica pode ter consequências imediatas como por exemplo a classificação, a estimativa e a previsão de possíveis resultados ou padrões de dados.

Os autores ainda evidenciam a ligação destes mecanismos com a colaboração humana ou *crowdsourcing* (quando uma entidade apela ao *input* de um grupo indefinido de utilizadores) para desenvolver futuras bibliotecas digitais multilingues ou sustentar as já existentes. (BUDZISE-WEAVER, CHEN e MITCHELL, 2011, p. 9) SORG (2011, p.2) propõe que métodos para explorar recursos multilinguísticos deveriam desenvolver-se dos mecanismos da Web 2.0 e explorar *Social Semantics*. *Social Semantics* define-se por, a partir de conteúdos gerados pelos utilizadores em ambientes na rede específicos que podem ser explorados, ser gerado conhecimento, esse conhecimento, ou enriquecimento da informação obtido é denominado por Social Semantics. Este método ofereceria vantagens relacionadas com a contribuição dos utilizadores através das redes sociais sobre qualquer tópico e em constante crescimento, de forma permanentemente atualizada.

CHEN e BAO (2009, p.3) afirmam que o estudo de utilizadores e do seu comportamento e opiniões pode ajudar a construir um sistema eficiente para o acesso à informação multilingue. Os sistemas mais tradicionais como ontologias muitas vezes focam-se apenas num domínio específico e normalmente trabalham com uma língua em específico. Tesouros muitas vezes não são atualizados frequentemente devido aos custos que tal revisão necessitaria. Tendo isso em conta é importante aproveitar as capacidades da Web 2.0. A Web 2.0 é definida por SORG como as atividades dos utilizadores na internet o que inclui todos os conteúdos gerados por estes. O autor ainda define *Social Semantics* como o conhecimento implícito gerado pela participação na Web dos utilizadores. A relação semântica destes dados que se apresentam de forma

agregada é definida através do uso coletivo. Isto pressupõe um acordo implícito para o entendimento consensual dos conceitos. (SORG, 2011, p.6) SORG (2011, p.7-8) ainda afirma que a Web 2.0 tem estruturas que permitem ser usadas para MLIR embora não possuindo ontologias, que contém o conteúdo gerado pelos utilizadores hierarquicamente organizado, intitulado de *category systems* que se definem como conceitos hierarquicamente organizados. Estes conceitos por sua vez são determinados por uma descrição textual em várias línguas. As categorias são, no fundo, unidades organizacionais que por sua vez podem ser definidas como *clusters* de conceitos relacionais e subcategorias que permitem uma classificação mais pormenorizada. As relações hierárquicas anteriormente referidas estendem-se entre as categorias. Os conjuntos de dados criados no âmbito da Web 2.0 são conhecimento agregado que pode ser considerado como *Social Semantics*. Estes sistemas não se encontram limitados nem pela quantidade de línguas, nem pelos domínios tratados, e encontra-se sempre atualizado. Desta forma são ideais para serem explorados em cenários de MLIR. Estes sistemas categorizados, compostos por estruturas semânticas, conseguem aproveitar este conhecimento gerado, pois a maioria das funcionalidades da Web 2.0 não incluem ontologias. As ontologias permitem a recuperação específica de um determinado domínio num cenário mais controlado, criando ligações e definindo cada objeto. Estes sistemas categoriais podem ser outra forma de explorar eficazmente e aproveitar fontes de informação criadas no ambiente da Web 2.0.

Pode ainda ser referida outra problemática da representação – a apresentação. A apresentação dos conteúdos, sendo que esta necessita que os caracteres pretendidos sejam representados. Para que isso se torne viável é necessária uma codificação normalizada para ser legível e para permitir o processamento textual como por exemplo a indexação pesquisável. Este problema já não se apresenta relevante para as línguas baseadas em alfabetos latinos mas é no entanto um desafio para outros alfabetos. Nem todos os sistemas de codificação permitem uma interoperabilidade eficaz e não existe um sistema de codificação que consiga lidar com todas as línguas. Outra problemática é o desenvolvimento de bibliotecas digitais multilingues pois esse tipo de bibliotecas para serem sustentáveis necessitam de equipas interculturais/transculturais para terem sucesso. (DIEKEMA, 2012, p.7) Registos bibliográficos multilingues são as informações constantes de um registo no âmbito de um catálogo que é concedido em várias línguas. Tal função integrada numa base de dados de biblioteca será benéfica para

o utilizador. Aqui levantam-se por exemplo problemáticas que envolvem a criação de novos sistemas que englobem essa funcionalidade e a duplicação ou mesmo a extensão de registos já existentes. (MAKARA, 2010, p. 3)

1.5. Critérios de avaliação para bibliotecas digitais multilingues

A avaliação pode definir-se como a determinação de desempenho ou de valor de uma determinada entidade. Existem diversas formas de avaliação e diferentes abordagens sempre dependendo dos objetivos, questões e abordagens metodológicas escolhidas. (SARAVECIC, 2009, p.5) Como VULLO (2010, p. 172-173) afirma, os modelos de avaliação de bibliotecas digitais podem seguir vários tipos de abordagens. A abordagem pode ser centrada no conteúdo, ou seja, a biblioteca digital é observada como um conjunto de dados e metadados. Outra abordagem é basear a avaliação em questões técnicas, ou seja, tendo em conta o sistema de *software*, ainda pode ser abordada da perspetiva dos serviços, neste caso observa-se a biblioteca digital como entidade que fornece um conjunto de bens intangíveis ou, por fim, uma abordagem baseada no utilizador, tendo aqui em conta o âmbito social e personalizado no qual ela se insere. SARAVECIC (2009, p.5) ainda destaca que a avaliação de biblioteca digital não necessita de envolver todo o universo da biblioteca digital, pode também centrar-se em certos aspetos e funcionalidades. Segundo este autor é ainda desconhecido como se pode avaliar uma biblioteca digital em toda a sua dimensão. Já ZHANG (2007, p.99-102) propõe um modelo holístico de avaliação de bibliotecas digitais. Neste modelo existem seis conjuntos, sendo que cada conjunto inclui vários critérios. Trata-se de um modelo de círculos concêntricos. O círculo superior destaca um ponto fulcral da biblioteca digital, o contexto da mesma. Na parte inferior estão destacados dois componentes fundamentais para o seu funcionamento, denominados pelo autor como conteúdo e tecnologia. O interface, para o qual todos os conjuntos contribuem, assume a posição central. Os círculos dedicados aos utilizadores e aos serviços apresentam-se do lado esquerdo e direito, estes representam a relação com o público e a envolvência com os agentes. Os três conjuntos que neste estudo se apresentam como mais relevante são o interface, os serviços e os utilizadores e por isso serão especificados. Os critérios de avaliação que podem ser destacados em relação ao Interface são: a facilidade de utilização, a eficácia e a consistência. Em termos de serviços ZHANG (2007, p. 99-102) refere critérios como, a acessibilidade, a fiabilidade, a capacidade de resposta, a utilidade para o público-alvo e ainda a integridade do sistema do acesso e da

recuperação da informação. Mais especificamente para as bibliotecas digitais ainda são destacados fatores como, a satisfação, o custo e o controlo no manuseamento por parte do utilizador. No que diz respeito aos critérios de avaliação no âmbito do utilizador, os que mais se destacam são: a precisão e o sucesso com a qual uma tarefa é executada, a satisfação e a eficiência.

DIEKEMA (2012, p.7-8 apud BIA et. al., 2005) utiliza os seguintes critérios relevantes para as bibliotecas digitais multilingues: Em primeiro lugar a gestão dos dados, os diferentes vocabulários, metadados e a indexação. (DIEKEMA, 2012, p.7-8 apud BIA, [et. al.], 2005) Em segundo lugar a representação onde são inseridos aspetos de apresentação. Ainda se refere as funcionalidades de interoperabilidade que são relacionadas com o interface da Biblioteca digital. MAEDA (2002, p.1) considera que as técnicas mais relevantes para contribuir para os problemas levantados pelo multilinguismo são as funcionalidades de visualização e de introdução de dados para texto multilingue, um algoritmo para a identificação automática de línguas e sistema de codificação de caracteres de documentos e por último uma técnica de recuperação de informação *cross-language* que se adequa a documentos de vários domínios. ARCHANA e DEVI (2011, p.27) afirmam que atualmente os sistemas de RI incluem classificação de documentos, categorização, motores de pesquisa, interfaces, visualização de dados, filtros de informação, processamento das línguas naturais ou linguagem de *query* e arquitetura de sistemas. Do ponto de vista dos recursos digitais inclui ainda *text mining*, ou seja, a obtenção de informação de elementos textuais, e recuperação multimédia. O autor ainda distingue num sistema de recuperação de informação entre pesquisa por navegação por envolvimento de hiperligações e pesquisa por *query* como por exemplo o caso do Google. Estes componentes aqui destacados apesar de não se focarem especificamente na avaliação representam fatores importantes a ter em conta na análise, pois consideram fatores que podem ser integrados nos aspetos fulcrais a serem analisados.

Ainda para selecionar alguns critérios é relevante explorar o projeto de investigação MultiMatch que estabelece alguns critérios para o desenvolvimento de um motor de pesquisa multilingue específico para acesso, organização e representação personalizada da informação de património cultural. Neste domínio existem alguns desafios em termos de sistemas de RI, como sejam o *web crawling* (um método de indexação dos conteúdos disponibilizados na internet, desta forma podem manter-se os

conteúdos, as bases de dados e as funcionalidades atualizados), acesso multilinguístico, acesso multimédia, processamento semântico e características da apresentação. Os resultados desejados em termos de funcionalidades, determinados para um tal sistema multilingue seriam em primeiro lugar evoluir de um *Focussed search engine* (motores de pesquisa verticais, que permitem a pesquisa em domínios específicos e permitem uma resposta à query mais sofisticada e com o objetivo de conseguir a recuperação de objetos complexos) para um motor de pesquisa multilingue que permite a apresentação, acesso, organização e personalização da informação disponibilizada. Esta evolução será conseguida através de *focused crawling* (forma de retirar informação centrada no tópico e de qualidade através de páginas web) e enriquecimento semântico que explora os metadados criados para o património cultural. O segundo resultado desejável é a indexação multilingue/multimédia. O objetivo não é o sistema apresentar documentos isolados mas expor resultados de pesquisa complexos que apresentam vários tipos de media e línguas em contexto, o que representa um desafio para a indexação. Ainda é considerada importante a extração e a classificação da informação, o que irá permitir aos utilizadores interpretar o valor da informação do património cultural ao representar objetos, não como itens individuais, mas como hiperligações para contextos relacionados. É também importante destacar a RI multilingue e multimédia onde são identificados problemas com matérias dependentes de língua (texto e áudio) e material independente de língua como vídeos ou imagens, nomeadamente no que diz respeito ao processo de pesquisa. Outro resultado é a interação centrada no utilizador, onde serão integradas técnicas automatizadas para a classificação automática de conceitos e será possibilitado ao utilizador formular a sua *query* adequadamente, utilizando qualquer língua pretendida. Valores como a flexibilidade no acesso aos conteúdos são também valorizados. (AMATO; [et. al.], 2007) FRANKLIN, KYRILLIDOU e PLUM (2009, p.35) referem que a usabilidade dos serviços e a sua avaliação formativa como determinante pois no ambiente de uma biblioteca digital existem diferentes interfaces e sistemas. É importante o estudo da usabilidade para fazer alterações aos sítios da internet e a apresentação dos conteúdos. Ainda afirma que este facto contribui para o poder competitivo da própria biblioteca digital pois existe melhoria nos recursos e nos serviços e pode também melhorar os resultados do ponto de vista do utilizador.

GÄDE (2011, p.2) distingue vários aspetos de acesso à informação em bibliotecas digitais multilingues. O primeiro é o Interface do utilizador multilingue que

a autora considera o aspeto mais elementar. Para a determinação da língua são utilizados dois métodos, no primeiro o utilizador pode seleccionar a língua do interface por um menu. Na segunda opção as definições de língua são automaticamente alinhadas com as definições do *browser* ou a localização geográficas (através do endereço de IP). O segundo aspeto destacado é a pesquisa multilingue que é considerada a componente mais fundamental. Existem três formas de implementar *Multilingual Information Access* (MLIA). A primeira por tradução da *query* (na qual a *query* é traduzida noutra língua que os documentos da coleção da biblioteca digital contêm), a segunda envolve a tradução do documento (os documentos da coleção são traduzidos para a língua da *query*) ou interlíngua (no qual a *query* e os documentos são traduzidos para uma língua específica que acaba pelo ponto de vista do utilizador por ser uma pesquisa monolingue). O terceiro aspeto distinguido por GÄDE (2011, p.3) é a filtragem e a representação dos resultados multilinguísticos. A representação pode ser efetuada de duas formas: ou através dos metadados ou ao nível do objeto digital. No caso de recursos textuais é necessário definir se as funções de tradução se aplicam ao documento ou aos metadados. Já dentro dos metadados a autora afirma que os candidatos mais importantes para a tradução são os títulos e as palavras-chave em assunto. Em termos de filtragem dos resultados estes podem ser implementados de duas formas ou por pesquisa avançada onde podem ser determinadas as línguas ou filtragem do conjunto de resultados após a primeira pesquisa com a possibilidade de seleccionar uma determinada língua.

Diretamente relacionado com aspetos linguística está o sistema de avaliação HeMT (Human Evaluation of Machine Translation) para avaliação a tradução por máquina de sistemas de recuperação de informação. Esta avaliação permite o crowdsourcing em traduções manuais e por máquina. Esta forma de avaliação foi criada para ser aproveitada por três tipos de utilizadores: tradutores, avaliadores e revisores. Este sistema limita-se no entanto a estudar os resultados de tradução a partir de um grupo de avaliadores, não utilizadores normais. Neste caso de avaliação humana de tradução por máquina (*Machine translation*) a vertente avaliada foi a tradução dos metadados presentes no sistema, sendo que foram escolhidos apenas 6 campos de metadados,: “coverage, creator, description, publisher, subject e title”. Os valores que foram atribuídos neste estudo foram a adequabilidade e a fluência. (CHEN, [et al.], 2012, p.2, p.6) Este estudo embora interessante por se focar na especificidade dos

aspectos de tradução e linguísticos de um sistema de informação e contribuir para melhores resultados por parte dos utilizadores, podia envolver o utilizador na sua avaliação.

Para concluir este ponto da revisão de literatura e referir um exemplo actual sobre a avaliação de bibliotecas digitais, mais especificamente o acesso e a recuperação da informação no âmbito do multilinguismo pretende-se especificar o seguinte caso. Trata-se de uma tese recente que apresenta um número significativo de categorias e critérios para a avaliação de bibliotecas digitais, este caso é referido e destaca-se por considerar na sua avaliação funcionalidades multilingues. (BARVE, 2012, p.73-96) elaborou várias categorias e critérios para avaliar bibliotecas digitais. No entanto após uma leitura e apreciação de todos esses critérios, e embora esses critérios estejam bastante completos, verificou-se que as funcionalidades multilingues poderiam ter mais visibilidade, tendo em conta a sua importância central. No ponto 3.9. do estudo de BARVE (2012, p.) com o título “Metadata Submission and Support” são destacadas as ontologias e os tesouros, que são sem dúvida uma componente central no que se refere ao acesso e à recuperação da informação. No entanto não são especificadas funções multilingues neste âmbito. Apenas é destacado que é necessário avaliar se existe ou não e qual o tipo de protocolo e de normas aos quais essas ferramentas linguísticas correspondem. Nos critérios relacionados com a área destacada como “Information Search & Retrieval”, mais especificamente no que diz respeito ao “browsing” não é referido qualquer papel relevante que funcionalidades multilingues possam ter neste âmbito. Ainda nesta mesma categoria, considerando a parte referente a pesquisa, “Search”, já existe o critério “Support multilingual search and retrieval” mas apenas na perspectiva de verificar se poderá existir ou não, não definindo qualquer impacto mais directo. Encontra-se, igualmente referenciada a importância do multilinguismo na categoria de “User Friendly Interface” na qual é dada importância ao facto do utilizador poder especificar a sua língua. Nesta categoria foram encontrados dois critérios interessantes assim como: “Can user interfaces be localized in their language?” e “Does the system provide multilingual access support?”. Outra questão interessante é o destaque dado ao “Translation Service” na categoria de “Advanced Services” que contribui não só para facilitar a pesquisa e a navegação mas é também considerada ao nível do documento. No entanto o que não se encontra integrado na sua definição é a

importância que, como já referido, as ferramentas disponibilizadas pela Web 2.0, possam ter neste campo de ação.

É necessário para o sucesso de uma biblioteca digital dar importância a componente da avaliação. Como CASTAGNÉ (2010, p.5) afirma ainda é verificada uma lacuna grande entre a teoria e a prática da avaliação. Trata-se de um campo de estudo que ainda se encontra em desenvolvimento e em crescimento constante, alcançando, aos poucos, a sua maturidade.

2. Abordagem metodológica

A metodologia escolhida no âmbito desta dissertação centra-se numa abordagem qualitativa, o tipo de estudo é descritivo sendo um estudo de caso. Neste âmbito o objeto do estudo de caso situa-se entre as organizações/instituições, mais especificamente, as bibliotecas digitais multilingues.

A questão de investigação colocada neste campo de ação é “Quais são as formas e as funcionalidades de acesso e recuperação de informação multilingue aplicadas no âmbito das bibliotecas digitais multilingues e como é que podem ser definidas formas de análise?”

A metodologia é centrada no estudo de caso para observar como num contexto real, no caso de uma biblioteca digital multilingue, se estabelece o processo de acesso e recuperação da informação. O acesso e a recuperação da informação é considerado, no âmbito desta tese, como um processo único que se irá dividir em várias fases, dentro das fases serão estabelecidos, através da revisão da literatura e da observação exploratória, critérios de avaliação. Por sua vez aos critérios de avaliação serão aplicadas através indicadores que se consideram pertinentes para a análise da fase do processo em questão.

Aspetos essenciais que devem ser respeitadas na recolha de dados são, a definição do que vai ser avaliado; quais os objetivos; e a especificação dos critérios. (CASTAGNÉ, 2010, p.1) Para este efeito foi feita uma primeira fase de análise sob a forma de observação exploratória. Como a realidade dos estudos de caso em toda a sua dimensão ainda não tinha sido observada decidiu-se proceder a uma primeira fase de observação informal.

Nesta observação adaptou-se a metodologia de avaliação *cognitive walk-through* na qual se assumiu o papel do utilizador com o objetivo de observação do interface para completar tarefas de pesquisa. (BARVE, 2012, p.63) O objetivo é a verificação de situações que possam causar desconforto da perspectiva do utilizador. Esse método foi aplicado aos três casos. Na técnica de avaliação utilizada, *cognitive walk-through*, o avaliador assume o papel do utilizador para submeter a testes o processo de acesso e recuperação de informação. Pretende-se desta forma uma simulação dos processos pelos quais o utilizador passa por todos os interfaces disponibilizados por uma biblioteca

digital. Esta técnica de avaliação tem revelado sucesso no que diz respeito à identificação de problemas, especialmente no que diz respeito às questões de terminologia ou outras tipologias de erro. Também a visibilidade de certas funções pode ser avaliada desta forma. (BARVE, 2012, p.63)

Decidiu-se após esta observação proceder a algumas mudanças nos critérios de análise criados, a partir da revisão de literatura, com o objetivo de adaptá-los melhor às necessidades específicas dos estudos de caso e completá-los com uma perspectiva mais direta de observação. Os critérios de avaliação foram agrupados em cinco grelhas de avaliação que serão explicadas em pormenor no ponto 3.2., criando assim uma sequência lógica e demonstrando primeiro o enquadramento teórico, no ponto 2.5., nos quais estas grelhas se baseiam. O resultado foi a criação de cinco grelhas de análise distintas. Tal justifica-se por existirem e serem destacadas durante a fase exploratória de observação de quatro grandes dimensões, consideradas para serem avaliadas dentro do campo do acesso e da recuperação da informação. A primeira dimensão é o interface de acesso, ou seja, a página inicial, que se justifica por ser o contacto inicial para aceder à informação onde deve ser imediato o acesso a funcionalidades multilíngues para o utilizador, estando estas claramente destacadas. A segunda dimensão são as várias funcionalidades de pesquisa e navegação que apresenta a fase intermédia, na qual fatores como a coerência das funcionalidades multilíngues podem ser melhor avaliados. A terceira dimensão é a apresentação de resultados e a filtragem dos mesmos. A quarta é centrada no contacto do utilizador com um recurso específico, denominado apresentação do recurso, ou seja, trata-se do interface no qual estão disponibilizados os metadados e toda a informação sobre um item particular. A divisão destas quatro dimensões de avaliação permite uma abordagem mais específica e aprofundada. Uma quinta dimensão considerada é dedicada a fatores externos, as ferramentas da Web 2.0, que através dos utilizadores pode melhorar e criar tradução de vários conteúdos.

As mudanças essenciais após a observação exploratória já referida estão relacionadas com os índices de pesquisa e a aplicação das funcionalidades de tradução. A este estudo ainda foram adicionados fatores como a flexibilidade e a coerência da recuperação da informação, que foram considerados essenciais após esta primeira avaliação, surgindo previamente direta ou indiretamente na revisão de literatura.

-Flexibilidade: Flexibilidade, neste âmbito, é avaliada pela existência em todos os interfaces de pesquisa de funcionalidades multilingues (ex: funcionalidade de tradução) para uma utilização flexível dos recursos por parte do utilizador.

-Coerência: A coerência é avaliada pelo número de línguas do interface e de pesquisa não difiram, de forma a prejudicar a pesquisa e que essas línguas sejam apresentadas com a mesma qualidade, sendo o critério de qualidade aqui definido como a apresentação sem incoerências, ou seja, completa, em qualquer língua.

SARAVECIC (2009, p.6) define a eficácia como a avaliação do desempenho de um sistema nas suas tarefas tendo em conta os seus objetivos definidos para o próprio sistema. Para isso é necessário especificar quais os objetivos de um determinado sistema. A eficiência está relacionada com aspetos financeiros, o trabalho realizado e o tempo investido.

Tendo esta definição em conta torna-se claro que é impossível, por uma questão de insuficiência de acesso a dados específicos, a avaliação da eficiência dos vários casos. No entanto a avaliação da eficácia, ou seja, avaliar se o desempenho das funcionalidades multilingues, que se propõem avaliar, de um determinado sistema, estão alinhados com os objetivos dos vários casos, é um dos objetivos deste estudo. Por isso a avaliação da eficácia é que se destaca por ser passível de avaliação.

Outra importante questão que deve ser destacada aqui é o facto de numa biblioteca digital existirem diferentes opções de pesquisa e navegação, podendo existir diferentes funcionalidades multilingues integradas nessas formas de acesso e recuperação da informação. O objetivo é não ignorar essa realidade mas sim detetar as diferentes estratégias existentes em cada um dos estudos de caso. Todos os critérios escolhidos, com exceção do tratamento da informação de ordem mais técnica, foram baseados na perspetiva do utilizador como já referido anteriormente.

2.1. Estudos de caso: critérios de seleção

Antes de proceder a escolha dos estudos de caso foram considerados critérios de exclusão e de inclusão para poder reduzir o universo dos casos possíveis e escolher os casos que se enquadram nos parâmetros que se pretendem avaliar. A escolha dos estudos de caso e o estabelecer de critérios é determinante para o sucesso do estudo e é assim uma parte integrante da abordagem metodológica.

Critérios de inclusão e exclusão:

Para critérios de inclusão foram escolhidos os seguintes:

- Orientação a um público-alvo internacional
- Contribuição dos recursos de várias instituições culturais em várias línguas
- Projetos recentes, não anteriores ao ano de 2000
- Interface disponível pelo menos em três línguas da União Europeia, sendo duas delas as línguas de estudo propostas

Para critérios de exclusão foram selecionados:

- Projetos de curta duração, sem um número de recursos significativo

O foco na escolha dos estudos de caso residiu principalmente num público-alvo que se pretendia ser o mais internacional possível (existindo assim a necessidade de uma biblioteca digital multilingue). Desta forma, e pela participação frequente de bibliotecas nacionais nos vários projetos selecionados, os estudos de caso seriam de um interesse maior se tivessem como alvo um público mais abrangente, não se centrando em projetos suportados apenas nacionalmente mas sim com uma relevância a nível internacional.

Desta forma foram selecionados três casos de estudo, todos com importância internacional, que representam funcionalidades de recuperação de informação multilingue em mais de três línguas, empenhados no desenvolvimento de técnicas cada vez mais eficazes para o objetivo de servir um público internacional e disponibilizar o seu serviço o mais eficazmente possível.

Escolheu-se neste âmbito a Europeia que embora destinada ao público europeu, foi igualmente criada para dar a conhecer o património cultural europeu a um público internacional. A segunda biblioteca digital multilingue selecionada é a World Digital Library (WDL) cuja importância a nível mundial tem crescido, contando com uma participação muito relevante em termos continentais, concentrando-se igualmente no património cultural. A única biblioteca selecionada que se centra num público internacional mas com a especificidade de ser mais direcionado a utilizadores infantis é a Internacional Children's Digital Library (ICDL) que foi escolhida por se situar no campo de ação multilingue por excelência e por se destacar no âmbito das bibliotecas digitais multilingues internacionais.

Para avaliar os estudos de caso é necessário escolher algumas línguas nas quais se pretende uma observação mais concreta. Seleciona-se como principais línguas de trabalho o português, o inglês, o alemão, o espanhol e o francês. O português devido à sua importância no âmbito de língua oficial do país na qual a tese se encontra a ser desenvolvida. O inglês por ser uma língua pivô presente em qualquer âmbito da União Europeia. O alemão e o francês por serem igualmente línguas fortes da União Europeia e por permitirem aumentar em termos de abrangência a abordagem pretendida ao tema do multilinguismo. O espanhol é escolhido também por uma razão de abrangência e pela produção de recursos tanto da Europa como da América Latina.

2.2. Casos selecionados: caracterização

Como o estudo desta dissertação se centra especificamente no estudo de caso foi considerado necessário a definição de cada caso escolhido. Convém salientar fatores como os objetivos, desenvolvimentos e fatores relevantes de cada caso para completar o estudo e poder criar uma visão global. Contribui para uma possível avaliação comparativa entre os casos destacados.

2.2.1. O caso Europeana

O portal da *Europeana* já disponibiliza uma vasta gama de conteúdos patrimoniais. Para esta agregação de recursos é necessário superar vários desafios, tais como os tipos heterogêneos de *media* (imagens, som etc.) que requerem organização e apresentação ao mesmo tempo. Já a descrição por metadados e por objetos chega a este portal em esquemas, formatos e línguas de descrição diferentes. Quando se considera o aspeto linguístico é sempre necessário considerar os termos em língua natural como os vocabulários controlados. Através da disponibilização de serviços de acesso multilinguístico, os conteúdos devem estar acessíveis de forma igualitária para todos os utilizadores da plataforma, e isto independentemente da sua língua nativa e da língua dos recursos. Os principais objetivos que se pretendem com as ferramentas e a infraestrutura multilingue da *Europeana* são a avaliação das necessidades dos utilizadores multilinguísticos, o repositório de recursos de línguas da *Europeana* para mapeamento e tradução, o mapeamento multilinguístico de esquemas de metadados para assuntos e outros vocabulários controlados, os serviços ou módulos de tradução de *queries* para a estrutura de recuperação da informação e pesquisa, a avaliação de protótipos de tradução de *queries* e por último uma estratégia para a integração destas

ferramentas no sistema da *Europeana*. Deve ser tido em conta que nem todas as línguas presentes na plataforma *Europeana* estão desenvolvidas, em termos de tradução, a níveis semelhantes. (PETRAS, 2011, p. 5-7)

Para ultrapassar a barreira linguística, as bibliotecas digitais, necessitam de oferecer um acesso multilinguístico aos conteúdos em múltiplas dimensões, tanto do ponto de vista do sistema como do ponto de vista do conteúdo, tendo como foco o utilizador e os processos por ele desenvolvidos na pesquisa e na recuperação da informação (EUROPEANA, 2012, p.4) As estratégias para ultrapassar essa mesma barreira de acesso multilinguístico da *Europeana* dividem-se em cinco aspetos. Em primeiro lugar destaca-se o interface multilinguístico para o utilizador, o que inclui a tradução de todos os conteúdos estáticos. O segundo prende-se com o enriquecimento multilinguístico de metadados de objetos e o mapeamento de conhecimento monolingue para uma rede semântica multilingue. Em terceiro lugar a estratégia foca a pesquisa multilingue que pode ser desenvolvida a partir da tradução de *queries*, tradução de documentos ou uma abordagem interlinguística (as *queries* e os documentos são traduzidos para uma determinada língua). A tradução dos documentos em si apresenta-se como uma funcionalidade futura que de momento ainda não é uma possibilidade. É ainda de destacar que a tradução de *queries* é considerada o método mais frequente no enquadramento dos sistemas de informação. A quarta estratégia é a apresentação multilingue de resultados o que inclui o apresentação dos resultados de acordo com a filtração por língua exigida pelo utilizador, tal é alcançado ao traduzir o objeto ou os metadados do objeto em questão. O último ponto estratégico é a navegação multilingue que a *Europeana* disponibiliza através do acesso a sugestões de imagens que não acompanhadas de texto que corresponde à língua do interface. (PETRAS, 2011, p. 7-8)

Os desafios futuros que o projeto *Europeana* enfrenta e que influenciam a experiência do utilizador são vários. Em primeiro lugar a pesquisa multilingue relacionada com a tradução, nomeadamente a problemática do licenciamento e a disponibilização de recursos linguísticos nas línguas necessárias. Outra problemática destacada é a tradução das *queries*, pois esta atualiza-se muito lentamente e a sua qualidade ainda é relativamente baixa. A *Europeana* tem considerado ao longo do seu projeto o enriquecimento da indexação de versões de metadados paralelos linguísticos mas este processo ainda se encontra em melhoria. Ainda funcionalidades de pesquisa para outros dispositivos como dispositivos móveis têm vindo a ser ponderados. A

utilização da colaboração de utilizadores tem sido outro ponto que tem vindo a ser estudada devido à especificidade dos termos utilizados em casos de objetos específicos. (PETRAS, 2011, p.17) Têm vindo a ser considerados no âmbito da *Europeana* modelos de interação com os utilizadores visto estarem a tornar-se numa área de investigação popular. Tal modelo irá ser desenvolvido para o acesso a esta plataforma com o objetivo de enriquecer o seu carácter multilinguístico e facilitar o acesso. Para esse efeito serão estudados os métodos de *user-assisted query translation* e *crowdsourcing* que apelam à interação ativa dos utilizadores com a plataforma. (EUROPEANA, 2012, p.7)

2.2.2. O caso World Digital Library (WDL)

A WDL foi criada pela UNESCO e outras 32 instituições e foi lançada em 2009. O objetivo é disponibilizar conteúdos de bibliotecas e arquivos a todo o mundo. As quatro principais razões para a existência deste projeto são a criação, a preservação, a partilha e a aplicação do conhecimento. (ABID, 2009, p.1) A WDL contempla a participação universal de todas as instituições culturais. No ano de 2012 já se contava com um total de 89 instituições parceiras em 45 países (entre elas bibliotecas, arquivos e museus). (WORLD DIGITAL LIBRARY, 2012, p.1) A WDL disponibiliza em linha, grátis e multilingue, materiais de vários países e culturas a nível mundial. Os principais objetivos da WDL são promover a compreensão internacional e intercultural, expandir o número de conteúdos culturais na internet, e permitir o acesso ao público em geral e, em particular, a académicos e professores. Os recursos no âmbito da WDL podem ser pesquisados por lugar, tempo, tipo de formato, e pela instituição contribuinte, tudo em várias línguas. Outras das funcionalidades são os *clusters* interativos geográficos, a *timeline*, a visualização de imagens avançada e a capacidade interpretativa. As línguas disponibilizadas para a navegação e descrição de conteúdo são o árabe, o chinês, o inglês, o francês, o russo, o português e o espanhol. Existem ainda mais línguas nos recursos providenciados nas suas línguas nativas. As funcionalidades de pesquisa e navegação facilitam a exploração do sítio da internet, tanto através de conteúdos temporais como culturais. (ABID, 2009, p.3-4) As prioridades a longo prazo da WDL, decididas em 2012, contam com o aumento do número de línguas nas quais o interface está disponível, adicionando pelo menos o hindu, o japonês e o alemão. Outra prioridade é o desenvolvimento de aplicações móveis. A terceira prioridade passa pela criação de áreas de foco temáticas em relação a acontecimentos históricos específicos ou áreas específicas culturais. (WORLD DIGITAL LIBRARY, 2012, p.4)

Segundo ABID (2009, p.5) as áreas principais nas quais a WDL se pretende destacar são: metadados consistentes, desta forma cada item é descrito por um conjunto consistente de metadados, ou seja, informação bibliográfica. Este conjunto inclui por exemplo informação geográfica, temporal e de assunto. O objetivo é criar ligações entre os recursos e permitir a motores de pesquisa externos aceder a essa informação. A descrição é outra área de destaque, pois cada item possui uma descrição escrita por especialistas fornecendo o contexto para o utilizador. O multilinguismo é igualmente uma aposta, assim, os metadados, a navegação e o conteúdo de suporte dos recursos estão traduzidos nas sete línguas das bibliotecas que contribuem para o seu carácter universal. A quarta área de importância é o desenvolvimento técnico da biblioteca digital que contribui para fatores como a catalogação ou funcionalidades multilingues. Neste âmbito foram desenvolvidos diversos serviços, como uma nova aplicação de catalogação que suporta os requisitos dos metadados, uma ferramenta centralizada com uma memória de tradução (para o trabalho dos tradutores), um interface que desperta a curiosidade para a exploração da página e a aposta em novas tecnologias que permite a redução do tempo de disponibilização dos conteúdos. A quinta área na qual a WDL aposta é a de uma rede colaborativa, que visa a abertura tanto no que diz respeito ao acesso aos conteúdos e parceiros, bem como a participação dos utilizadores. A WDL tem o objetivo de agregar os conteúdos de e sobre todos os países do mundo. Este projeto pretende igualmente ajudar bibliotecas em todo o mundo na digitalização de parte das suas coleções, de forma a beneficiar a área da investigação das ciências sociais. (OUDENAREN, 2009)

2.2.3. O caso International Children's Digital Library (ICDL)

O projeto ICDL teve o seu início em 2002 e foi criado pela Universidade de Maryland e pelo Arquivo da internet, com a ajuda e financiamento da Fundação Nacional da Ciência (NSF). Os objetivos do projeto prendem-se com a criação de uma coleção na qual se inserem 10 000 livros para crianças em cem línguas. Colaborar com crianças para melhorar o *design* e desenvolver novos interfaces para a pesquisa, navegação, partilha de leitura de livros de forma online e ainda a avaliação do impacto de materiais multiculturais, tanto no âmbito do público infantil, como no caso de escolas e outras bibliotecas. (HUTCHINSON; [et al.], 2006, p.5) A ICDL foi fundada pela Fundação Nacional de Ciência, o Instituto de Serviços de Museus e Bibliotecas (IMLS) e investigação Microsoft. Trata-se de um projeto de investigação do laboratório de

interação humana e computadores e da Escola de Estudos de Informação da Universidade de Maryland. (ICDL, 2010?)

A ICDL pretende quebrar as barreiras linguísticas ao permitir aos utilizadores ler livros tanto do seu país como, simultaneamente, expô-los a diferentes culturas. Desta forma, esta biblioteca digital pretende criar uma comunidade centrada na aprendizagem. (HALL, 2010, p.2) É considerado determinante tornar público, os conteúdos de bibliotecas digitais em linha para todos os utilizadores de forma global. Não apenas para disseminação cultural e consciência cultural, mas igualmente para dar mais poder à comunidade biblioteconómica. A ICDL é um projeto ambicioso em termos de pesquisa, pois foi criada para um público abrangente, o público-alvo da biblioteca abrange, potencialmente, todas as culturas e línguas do mundo. Ainda é necessário destacar que o conteúdo da biblioteca não está localizado. Um dos objetivos da biblioteca é fornecer aos utilizadores livros de diferentes culturas e línguas. Consequentemente, o interface desta biblioteca suporta várias línguas e culturas de forma simultânea. A ICDL ainda se destina a uma grande variedade de grupos etários tendo tanto utilizadores dos 3 aos 13 anos como utilizadores adultos que trabalham com crianças. (HUTCHINSON; [et al.], 2006, p. 4-5) O objetivo principal de ICDL é seleccionar, colecionar, digitalizar e organizar materiais de leitura infantis nas línguas originais e criar formas apropriadas de acesso para os utilizadores. Um dos maiores benefícios é a variedade de línguas nas quais os livros estão escritos. Existem mais de 50 línguas nas quais existe, pelo menos, um livro. A apresentação de tantas opções em termos de línguas reforça o seu carácter internacional. (HALL, 2010, p.4-6) A ICDL em termos de pesquisa possui quatro opções: a pesquisa simples, a pesquisa avançada, a pesquisa por localização e, finalmente, por palavra-chave. Na pesquisa simples é apresentada uma lista de 17 línguas para os conteúdos nesse específico ecrã de pesquisa. A pesquisa avançada permite a pesquisa por público, aparência, conteúdos, tipo e assunto. O público consegue ser especificado por idade, língua e data de publicação. Dentro da categoria tipo estão inseridos por exemplo classificações e géneros. No caso da pesquisa por localização trata-se de um globo interativo que permite a exploração. Por último a pesquisa por palavra-chave é uma opção simples que inclui uma lista com 54 línguas. Outras funcionalidades que convém considerar são as funcionalidades de navegação por lista de livros, autor, ilustrador, aquisições recentes, livros premiados, atividades e exposições. (BUDZISE-WEAVER, 2011, p.7) Em Novembro de 2002, no início do

projeto, foi apresentada a versão inglesa da biblioteca. Já em 2004, os metadados da biblioteca foram traduzidos nas línguas nativas permitindo aos utilizadores escolherem a língua dos metadados dos vários conteúdos. Em 2006 foi iniciada a fase final do projeto que envolveu a tradução do interface em diferentes línguas e o ajustamento do *design* visual do mesmo tendo em conta as normas culturais de cada língua. (HUTCHINSON; [et al.], 2006, p.5)

2.3. Critérios de análise

A escolha dos critérios de análise apresentados neste ponto resultou da identificação de quatro fases distintas do processo de acesso e da recuperação da informação, verificadas em cada um dos casos. Desta forma, foram atribuídos critérios que se centram principalmente, tendo em conta os objetivos desta dissertação, em funcionalidades de pesquisa e navegação. É igualmente considerada a questão de como o multilinguismo se afirma nestes campos. Esta afirmação centra-se principalmente nos conceitos de tradução e na pesquisa e navegação multilingue (considerando também os índices de pesquisa como tendo um papel fundamental) que são considerados essenciais, como já se verificou na abordagem teórica, para o utilizador aceder de forma conveniente à informação. Sem estes fatores é impossível uma biblioteca digital considerar-se multilingue. Tendo isso em conta foram estes os aspetos que influenciaram mais a escolha dos indicadores neste estudo.

Os critérios de análise foram organizados em cinco grelhas de análise as quais serão utilizadas para analisar os estudos de caso de forma textual mas organizada na linha de raciocínio que os critérios apontam. A disposição dos critérios em grelhas de avaliação permite uma observação mais detalhada de cada fase do processo do acesso e da recuperação da informação.

a) Interface – Primeira fase do processo de acesso e recuperação da informação

Nesta categoria o interface define-se como a página principal de acesso, ou seja, a página inicial. Considera-se este aspeto o primeiro passo do acesso de recuperação da informação da perspetiva do utilizador. É devido a isso considerado o mais relevante, como já foi referido, pois determina em primeiro lugar um fator de ambientação para a interação. Este primeiro passo do processo necessita de estar coerente com o resto do

processo sendo que a coerência é uma característica determinante em todo o processo como será explicado de forma mais aprofundada posteriormente.

Tabela I - Interface

Fase de acesso	Interface	Visibilidade	Visibilidade das funções multilíngue -utilização intuitiva da perspectiva do utilizador
		Línguas disponibilizadas	-Número de línguas nas quais se encontra o interface -Relevância das línguas tendo em conta o âmbito e os objetivos da biblioteca digital multilíngue
		Funcionalidades de tradução	Tradução de todos os conteúdos textuais presentes no interface -Relevância dos conteúdos não traduzidos

b) Pesquisa e navegação

Esta é considerada a fase mais complexa em todo o processo do acesso e da recuperação da informação. Diferentes formas de pesquisas e de navegação podem ser consideradas no âmbito de uma biblioteca digital, ainda podem ser abordadas funcionalidades de pesquisas e navegação que se adaptam especificamente a uma biblioteca digital multilíngue. Esse tipo de bibliotecas pretendem ter abordagens o mais globais possíveis, o que se deve refletir também nos tipos de pesquisa e navegação fornecidas ao utilizador.

Tabela II – Pesquisa e navegação

Fase de RI	Pesquisa por <i>query</i>	Funcionalidades de pesquisa	Quais as funcionalidades de pesquisa -Existência de funcionalidades especificamente multilíngues
		Línguas disponibilizadas	Línguas disponíveis nas funcionalidades de pesquisa -Número de línguas disponibilizadas por funcionalidade -Coerência das línguas disponibilizadas com as línguas disponibilizadas na página principal
		Funcionalidades de tradução	Tradução dos elementos textuais presentes nas várias funcionalidades de pesquisa -Relevância dos elementos não traduzidos
	Navegação	Funcionalidades de navegação	Existência de funcionalidades específicas que permitem a navegação multilíngue
		Línguas disponibilizadas	Línguas disponíveis nas funcionalidades de navegação -Número de línguas disponibilizadas por

			funcionalidade -Coerência das línguas disponibilizadas com as línguas disponibilizadas na página principal
		Funcionalidades de tradução	Tradução dos elementos textuais presentes nas várias funcionalidades de pesquisa -Relevância dos elementos não traduzidos

c) Fase de representação e filtragem dos resultados

Esta grelha de análise trata de uma das últimas fases do processo do acesso e da recuperação de informação. Nesta grelha decidiu-se aliar duas fases do processo, a fase da representação dos resultados e a posterior fase de filtragem dos mesmos. Juntaram-se estas duas fases por estarem intrinsecamente ligadas e desta forma influenciarem-se, não sendo independentes. No momento da representação dos resultados o utilizador pode escolher o recurso desejado, ou então proceder à filtragem destes resultados para encontrar de forma mais refinada o que é pretendido. Como tal consideram-se essas duas fases indissociáveis. É apenas necessário considerar que as funcionalidades de filtragem, em alguns casos, podem ser diferentes das funcionalidades de pesquisa anteriores, tendo em consideração os índices de pesquisa. Estas podem variar pela necessidade de reduzir os resultados e podem ter critérios como a filtragem por língua do documento ou país de proveniência. Estes índices para a filtragem são os mais desejáveis para o acesso e a recuperação da informação multilingue, por se centrarem mais na perspetiva linguística e de proveniência.

Tabela III – Representação e filtragem dos resultados

Fase de representação	Representação dos resultados	Opção de língua	-Existem opções de escolha de língua no
-----------------------	------------------------------	-----------------	---

dos resultados			interface da representação dos resultados -Quantas línguas são disponibilizadas
		Funcionalidades de tradução	Tradução dos elementos textuais presentes na representação dos resultados -Relevância dos elementos não traduzidos
Fase de Refinamento	Filtragem dos resultados	Filtros de informação	Existe filtragem dos resultados da pesquisa por língua ou país?

Quando no âmbito desta grelha de análise se refere a relevância de elementos não traduzidos, essa relevância é apenas destacada de três formas: essencial, ou seja, é determinante a tradução do elemento para o processo de acesso e recuperação da informação; intermédia, quando pode influenciar a recuperação da informação mesmo não interferindo com as funcionalidades mais destacadas (pesquisa e navegação) ou secundário, ou seja, trata-se de um elemento que é considerado não ter influência no processo em questão.

d) Representação do recurso

Ao nível do recurso, ou seja, quando o que o utilizador escolheu um determinado item para responder à sua questão de pesquisa inicial, as funcionalidades multilingues são essenciais para garantir o acesso a informação. Para que no último momento as necessidades de informação sejam respondidas são necessárias a tradução de metadados e de elementos textuais. Ainda mais benéfica será a tradução do documento em si, quando este possui conteúdo textual. Tal como o interface esta é uma das fases essenciais e críticas no acesso da informação e onde deve existir grande investimento por parte dos criadores.

Optou-se por separar as duas valências de tradução. Em primeiro lugar, a tradução de metadados e em segundo lugar a tradução dos elementos textuais. A tradução de metadados corresponde apenas aos campos pré-estabelecidos, que dominam cada conceito no âmbito dos metadados definidos para cada sistema, ou seja, aplica-se apenas à tradução da denominação do campo dada em cada caso. São considerados elementos textuais todos os elementos presentes no sítio web, que, neste caso, influenciem ativamente o acesso e a recuperação da informação. São igualmente considerados elementos textuais aqueles que se encontram no campo de descrição dos metadados, ou seja, não a denominação do campo mas sim as particularidades de cada recurso que corresponde à informação pretendida em cada campo de metadados.

Tabela IV- Representação do recurso

Fase da recuperação do objeto	Tradução	Mecanismos de tradução	-Tradução por query -Tradução dos documentos digitalizados
		Tradução de Metadados	Tradução dos metadados para a língua na qual a pesquisa foi feita
		Tradução do texto	Tradução de todos os elementos textuais -Relevância dos elementos textuais não traduzidos

e) Ferramentas - Web 2.0

As funcionalidades da Web 2.0 tornaram-se relevante para qualquer biblioteca que se apresenta no âmbito digital. Estas ferramentas não são apenas para dar visibilidade ou para efeitos de marketing. Também podem contribuir, de forma cada vez mais eficiente, para o funcionamento eficiente de uma biblioteca digital. Desta forma é possível, a partir dos utilizadores, criar e enriquecer a informação já existente na biblioteca. Neste caso esse enriquecimento pode ser feito através da tradução de conteúdos. Tal pode ser feito de diversas formas, assim como, pedidos de colaboração no interface ou

crowdsourcing. É determinante avaliar a existência dessas funcionalidades no âmbito desta tese por, do ponto de vista do utilizador, se tratar de uma ferramenta essencial de criação de conhecimento e de interação com o próprio sistema que disponibiliza a informação.

Tabela V – Funcionalidades Web 2.0

Ferramentas adicionais	Web 2.0	Criação da informação	Existem formas de envolver o utilizador na melhoria e na realização de traduções a partir de ferramentas disponibilizadas no espaço da biblioteca digital?

3. Análise dos resultados

A fase da recolha de dados ocorreu entre os dias 14 de Julho e 14 de Agosto. Os estudos de caso foram analisados pela seguinte ordem: ICDL, WDL e Europeana. Os sítios web foram consultados sempre a partir do mesmo computador com o sistema operativo Windows Vista. O *browser* utilizado para aceder às bibliotecas digitais foi o Windows Internet Explorer 9.

Cada sítio Web é observado sobre o ponto de vista do utilizador sendo essencial analisar as funcionalidades de língua, especialmente no que diz respeito à recuperação de informação. Fatores como: quais as línguas disponíveis, existe coerência na representação, a clareza da representação e a tradução completa são pontos essenciais que foram abordados.

A apresentação e discussão dos resultados de dados será, em primeiro lugar, conforme a grelha de análise anteriormente apresentada. Seguir-se-ão os pontos de avaliação, referentes a cada estudo de caso. Após essa avaliação será analisado de forma exhaustiva cada um dos casos, serão discutidos os resultados obtidos, estabelecendo uma comparação entre os casos. O passo seguinte será uma discussão envolvendo todo o enquadramento teórico, discutindo os objetivos, os resultados obtidos e destacando os pontos fundamentais da dissertação.

3.1. Análise e recolha de dados dos estudos de caso

1 – Interface

a) ICDL

Logo na primeira fase de acesso - o interface - este está disponível em cinco línguas: inglês, espanhol, francês, língua mongol e russo. Em termos de visibilidade, a funcionalidade de tradução está destacada e é imediatamente reconhecível.

Tendo em conta que um dos objetivos destacados por esta biblioteca digital, como já referido, é “a criação de uma coleção na qual se inserem 10 000 livros para crianças, em cem línguas”, a existência de apenas cinco línguas no interface pode ser insuficiente, tendo em conta que o interface é a primeira barreira linguística na perspectiva do utilizador. Ainda mais insuficiente se apresenta quando um dos

principais públicos para este projeto são crianças. Nesse caso o interface apesar de apresentar línguas relevantes em termos internacionais, torna-se ainda mais insuficiente.

No que diz respeito às funcionalidades de tradução estas são avaliadas nas línguas que foram propostas. Tanto em inglês, como em espanhol e francês, as línguas que se pretendem avaliar mais concretamente, quando é selecionada a língua, o interface permanece idêntico estando tudo completamente traduzido, excetuando alguns elementos de relevância secundária. No primeiro interface torna-se imediatamente claro que existe uma forte aposta na navegação e não na pesquisa, sendo que não existe a possibilidade de inserir uma questão de pesquisa no interface inicial. De forma central existe um ícone que se denomina “read books” acompanhado de algumas possibilidades de pesquisa e navegação, sendo essas não diretamente acessíveis, nem dando indicação de quais são as línguas na qual a pesquisa é possível. (ANEXO I; II)

b) Europeia

O interface de acesso, ou seja, a página inicial da Europeia encontra-se disponível em inglês, catalão, alemão, búlgaro, basco, checo, dinamarquês, grego, espanhol, estoniano, francês, gaélico, islandês, italiano, lituano, letão, húngaro, maltês, holandês, norueguês, polaco, português, romeno, russo, língua eslovena, eslovaco, suómi, sueco, finlandês e ucraniano. A funcionalidade de tradução do interface é bastante visível e destacada.

Neste caso, verifica-se um forte interesse em representar no interface as línguas que têm mais interesse ao nível europeu, o que é compreensível devido à sua natureza como biblioteca digital gerada para representar o conhecimento e a riqueza cultural a nível europeu. No entanto, também se destina a representar e competir a nível internacional com outros projetos de bibliotecas digitais e devido a esse facto seria benéfico disponibilizar igualmente o interface para públicos mais internacionais, por exemplo em língua chinesa, que poderá representar uma grande comunidade de utilizadores a nível internacional. Porém, pode considerar-se que, em conformidade com os objetivos da biblioteca, estes são cumpridos com as línguas apresentadas no interface.

Ativando a funcionalidade de tradução do sítio web para português são traduzidos a maior parte dos elementos que compõe o interface. Um elemento

designado por “Featured Partners”, por exemplo, continua em inglês. O Item em destaque e as notícias do blogue também continuam a apresentar-se em inglês, não apresentando qualquer forma de tradução imediata, assim como as legendas das imagens que aparecem com as notícias em destaque imediatamente por baixo da pesquisa. Como já referido na revisão de literatura a Europeia apenas se compromete com a disponibilização da tradução dos conteúdos estáticos. Pode concluir-se que ainda continua a existir uma representação inglesa mais completa do que a das outras línguas. Logo, um dos objetivos da União Europeia e que a Europeia sustenta, a promoção da diversidade linguística na sociedade, pode estar aqui em risco pois é apoiada uma língua em específico. (PETRAS, 2011, p.5) Ativando a mesma funcionalidade na tradução do interface em alemão, francês e espanhol a apresentação continuam com as mesmas inconsistências. Pode então considerar-se que a relevância dos conteúdos não traduzidos é intermédia, pois apesar de não influenciar diretamente a funcionalidade de pesquisa exposta, pode influenciar a navegação dentro dos conteúdos novos e destacados. (ANEXOS III; IV)

c)WDL

O primeiro interface de acesso na WDL encontra-se disponível em sete línguas: Português, Inglês, Espanhol, Francês, Árabe, Chinês e Russo. A funcionalidade de tradução do interface está num ponto de grande visibilidade e ocupa um lugar central no interface.

Em termos de representatividade de línguas, obviamente ao designar-se “World Digital Library” e conter recursos internacionais, um universo global de línguas disponibilizadas seria igualmente desejável. No entanto, sem ignorar que se trata de um projeto em desenvolvimento, a aposta nas línguas escolhidas para este início é considerada representativa. Como já referido na revisão de literatura, o projeto aposta na futura expansão para o hindu, japonês e alemão para ganhar mais visibilidade e “promover a compreensão internacional e intercultural”.

É de destacar que todos os conteúdos do interface em qualquer língua escolhida são traduzidos de forma completa, tornando o interface bastante acessível e nunca existindo qualquer tipo de perda de informação. Embora a quantidade de línguas ainda possa estar a ser desenvolvida, em termos qualitativos, no que diz respeito à interface, existe uma forte aposta na igualdade de acesso para qualquer língua. (ANEXO V; VI)

2 - Pesquisa e navegação

2.1. - Pesquisa

a)ICDL

Pesquisa avançada

O interface da pesquisa avançada encontra-se disponível nas seguintes vinte línguas: inglês, árabe, chinês (simplificado e tradicional), croata, dari, filipino, francês, alemão, hebraico, italiano, coreano, afegão, língua persa, polaco, português, mongol, russo, espanhol e tailandês.

A língua da palavra-chave inserida na query também pode ser especificada. Entre as línguas determinantes para este estudo, o inglês, francês, alemão, português e espanhol, ainda existem mais 74 línguas disponibilizadas nesta funcionalidade. Este fator contribui obviamente para o carácter multilíngue da biblioteca, sendo uma funcionalidade que engloba imensas possibilidades para os utilizadores. No entanto, um ponto de crítica aqui possível é a discrepância entre as línguas disponibilizadas em funcionalidades e interfaces diferentes, o que pode causar alguma confusão por parte do utilizador. O interface de acesso é apresentado em apenas cinco línguas enquanto já na definição de língua da query estão disponíveis 79 línguas. Tal não contribui de todo para a coerência do processo de acesso e recuperação da informação. Embora se apresente com uma funcionalidade flexível, ou seja, sempre disponível mesmo nas várias componentes que se apresentam na pesquisa, destaca-se a pergunta de “como um utilizador pode chegar a esse ponto da pesquisa”. Um utilizador alemão que não trabalha com mais nenhuma língua passará do interface? Chegará ele a todo este universo complexo onde a sua língua apenas existe na fase de pesquisa?

Como já referido na revisão de literatura a ICDL é um projeto muito ambicioso no que diz respeito às funcionalidades de pesquisa. A pesquisa foi criada para um público multilíngue e multicultural. (HUTCHINSON; [et al.], 2006, p. 4-6) No entanto, a questão persiste se, com as incoerências já identificadas, isto se traduz num benefício real.

Outra crítica possível é a complexidade existente. Pode ser definida a língua da interface e da query, mas sendo uma parte do público-alvo um público infantil, a aposta

não deveria ser de simplificação destas funcionalidades multilíngues? Trabalhando até com ícones das bandeiras dos diversos países podia tornar esta funcionalidade mais simples.

Dentro da pesquisa avançada existe ainda um campo denominado por “Audiência” e dentro desse campo um índice denominado “Língua” que apresenta um total de 62 línguas e o número de conteúdos disponíveis para cada língua. Esta funcionalidade permite restringir os resultados da pesquisa a uma determinada audiência tendo em conta a sua língua. Trata-se neste caso de uma funcionalidade de pesquisa especificamente linguística. Ainda dentro do campo “Conteúdo” existe a possibilidade de restringir a pesquisa por continente ou país, estes índices comprovam novamente a preocupação com o multilinguismo e multiculturalismo. Obviamente que, mesmo nesta funcionalidade, apesar de ser uma aposta bastante útil para o utilizador, nesta situação põe-se novamente a questão da discrepância entre as línguas.

Em termos de funcionalidades de tradução, quando estas são aplicadas para as línguas destacadas para este estudo verifica-se que os títulos de alguns recursos apresentados no interface não são traduzidos. Bem como não é traduzida a funcionalidade de navegação que se apresenta na pesquisa avançada, “From our shelves”. Esta inconsistência é considerada de relevância intermédia pois, apesar de poder interferir no processo de acesso e recuperação da informação, não se tratam de conteúdos determinantes para a funcionalidade de pesquisa. (ANEXO VII)

Pesquisa Simples

Existe de novo a definição, como em quase todos os interfaces disponibilizados, da tradução de todo o interface, disponibilizado nas mesmas línguas que a pesquisa avançada. Nesta tipologia de pesquisa não existe nenhuma funcionalidade específica no âmbito do multilinguismo que permita influenciar os resultados. A única funcionalidade que existe é novamente poder escolher a língua da query, sendo que estão disponíveis as 79 línguas já referidas. Apresenta-se novamente a questão da pouca coerência desta funcionalidade do ponto de vista global de todo o processo.

Quando é acionada a funcionalidade de tradução podem verificar-se duas inconsistências: os elementos “Featured Books” e “From Our Shelves” não aparecem traduzidos na versão alemã e alguns títulos expostos dos livros também não têm

tradução. Já na versão francesa e espanhola existe a tradução integral, no entanto títulos de livros encontram-se ainda em inglês. A versão portuguesa apresenta as mesmas falhas que a alemã, estando também alguns elementos em inglês e nem todos os títulos traduzidos. Nenhuma língua, à exceção do inglês, tem a tradução de todos os elementos do interface e do título dos livros. Esta desigualdade não contribui para os objetivos das bibliotecas digitais multilingues, sendo que a lógica subjacente é o acesso igualitário dos utilizadores de várias culturas e línguas. Quando disponibilizada uma língua para os utilizadores, esta, no caso ideal, deve possuir uma tradução completa dos elementos que se inserem no ambiente pelo qual o utilizador navega.

Pesquisa por palavra-chave

Esta é a última funcionalidade de pesquisa analisada no âmbito da ICDL. Esta funcionalidade visa a pesquisa de uma palavra-chave e disponibiliza, como as funcionalidades anteriores, a seleção da língua da query novamente em 79 línguas. Como nos outros menus de pesquisa, existe a possibilidade de mudar a língua do interface. A tradução deste interface seja no caso alemão, francês, português ou espanhol é completa.

2.2. – Navegação

a)ICDL

Busca por local

Considerou-se esta funcionalidade, não uma funcionalidade de pesquisa mas de navegação, pelo facto de não ser inserida qualquer questão de pesquisa que possa ser considerada para efeitos de recuperação de informação. No interface desta funcionalidade existe a representação de um globo no qual se consegue navegar nos livros a partir da sua localização geográfica. Como a localização geográfica pode ser relacionada diretamente com o fator língua está é considerada uma funcionalidade específica de navegação multilingue direcionado especificamente a um público multicultural e multilingue, tal como esta biblioteca digital pretende.

As línguas disponibilizadas para a tradução do interface são idênticas às das funcionalidades de pesquisa - vinte línguas. Neste caso, em qualquer língua está

disponível uma tradução completa de todos os conteúdos e elementos que integram o interface.

Após selecionar-se um continente, neste caso, o continente sul-americano já não se encontra no interface e já não há qualquer funcionalidade multilingue. Nem nos vários índices de pesquisa foi verificada a funcionalidade, o que representa uma falha em termos de navegação multilingue podendo ser necessário em qualquer momento mudar a língua de navegação pretendida. Neste interface estão apresentados vários livros através de títulos de capa, sendo ainda indicado por baixo do título, em que língua é que o livro se encontra.

b)Europeana

Pesquisa

No caso da Europeana existe apenas destaque a funcionalidade de pesquisa que se apresenta logo no interface de acesso. Neste caso, os elementos não traduzidos quando se efetua a tradução do interface são os mesmos que já foram destacados anteriormente na página principal. A funcionalidade na pesquisa que mais se assemelha a uma funcionalidade específica de multilinguismo é a pesquisa por “lugar”. No entanto, torna-se óbvio que este índice de pesquisa é baseado, como os outros índices, na pesquisa consoante os campos de metadados, ou seja, trata-se de um método de pesquisa baseado em ponto de acesso específicos.

c)WDL

Pesquisa

À semelhança da Europeana existe apenas uma funcionalidade de pesquisa generalista logo na página principal. Tudo o que se referiu acerca das funcionalidades multilingues do interface verifica-se igualmente aqui. Não existe nenhuma função da pesquisa que se refere especificamente a questões relativas ao multilinguismo ou ao multiculturalismo.

Navegação

As funcionalidades de navegação situam-se igualmente no interface de acesso. Existe um mapa-mundo que permite a recuperação de recursos consoante a área

geográfica. Esta pode ser considerada uma funcionalidade multilíngue. Existe igualmente a possibilidade de navegar por lugar, sendo que esta funcionalidade parece em todo semelhante à funcionalidade já referida e apresenta até uma listagem dos países. As línguas disponibilizadas são sempre, mesmo nas possibilidades de navegação onde o interface já é outro, as mesmas da página principal. Não existe em qualquer página a possibilidade de escolher qualquer outra língua que não as do interface, ou seja, a língua da pesquisa é sempre a língua do interface.

3- Representação e filtragem dos resultados

a)ICDL

No interface da representação dos resultados existe a opção de escolha de língua igual aos outros interfaces. São disponibilizadas as vinte línguas. Tanto em português como em espanhol, francês e alemão, a tradução do interface é completa, mesmo o título dos livros de todos os resultados. As opções de filtragem na pesquisa são equivalentes às opções que já se podiam utilizar no momento de efetuar a pesquisa.

b) Europeana

As línguas disponibilizadas são em tudo iguais às disponibilizadas nos outros interfaces. No entanto, a denominação dos recursos não se encontra traduzida para nenhuma língua, sendo esta a inconsistência em termos de relevância considerada essencial, uma vez que dificultará sempre toda e qualquer pesquisa por parte do utilizador para encontrar o recurso pretendido. A filtragem dos resultados possui um filtro pela “língua de descrição” e pelo “país fornecedor dos conteúdos”. Estes filtros em termos de flexibilidade são bastante positivos e permitem a um utilizador multilíngue restringir os seus resultados segundo as suas necessidades linguísticas específicas. (ANEXO VIII)

c) WDL

No que respeita às línguas disponibilizadas, estas são as mesmas que estão presentes nos restantes interfaces. Porém, a língua permanece sempre a mesma quando se pretende traduzir uma pesquisa, ou seja, tudo o que diz respeito aos recursos recuperados não é traduzido. A pesquisa do termo pretendido tem que ser realizada de

novo. Esta funcionalidade poderia ser automática. No entanto se toda a pesquisa é realizada de novo a representação dos resultados é apresentada na língua pretendida.

Como se verificou existem filtros especificamente do foro multilingue, neste caso os filtros são denominados “lugar”, filtro que se destina a restringir a pesquisa pela área geográfica e “idioma” que permite escolher o idioma dos recursos pretendidos. Como já referido no caso da Europeana, esta possibilidade de filtragem dos resultados confere um carácter de flexibilidade à pesquisa. (ANEXO IX)

4 – Representação do recurso

a) ICDL

Quando se entra no domínio específico da representação de um determinado recurso não existe nenhuma funcionalidade que permite a tradução do interface inteiro. Existe uma função de tradução específica para o recurso. O número de línguas disponíveis para traduzir a informação de cada recurso varia de objeto para objeto. Desta forma um determinado recurso pode ter apenas três línguas disponíveis para a tradução (ANEXO X, XI) ou ir até nove línguas. Não existe qualquer tipo de coerência nesta apresentação, podendo existir imensas discrepâncias de objeto para objeto. A denominação dos campos de metadados permanece sempre em inglês, independentemente da língua. Os elementos textuais, ou seja, o conteúdo descritivo dos metadados, na sua grande maioria encontram-se traduzidos, assim como o título. A biblioteca afirma que a tradução de metadados já se encontra finalizada, no entanto verifica-se que apenas a componente textual foi alvo de tradução. Outros elementos assim como os termos “Author” e uma parte essencial chamando atenção para a possível leitura do livro “Read this Book” permanece em inglês, seja qual for a língua escolhida. HALL (2010, p.4-6) afirma que a apresentação de tantas opções de línguas reforça a existência da ICDL como biblioteca digital internacional e multilingue. Embora a intenção da integração dessas várias línguas seja nesse sentido, não é o que se verifica na prática, na qual fatores como a incoerência vêm destabilizar o acesso e a recuperação da informação. Para além de poder levar a confusão por parte do utilizador, não torna o interface claro e fácil de utilizar. Tendo em conta que se trata de uma biblioteca digital também destinada a um público infantil ainda se torna mais pertinente corrigir este tipo de fraquezas.

É também importante destacar que o texto dos livros não se encontra traduzido, nem existe qualquer opção para a tradução do conteúdo. A única possibilidade de encontrar um recurso em línguas diferentes para além da original do documento é quando a biblioteca digital tem digitalizado edições da mesma obra em várias línguas. Não se trata de uma funcionalidade mas sim de uma disponibilização do mesmo recurso em edições de várias línguas. No entanto pela observação foi possível verificar que isto não ocorre na maior parte dos casos.

b) Europeana

Em 2011 PETRAS (2011, p.5) afirma que a Europeana dá acesso a mais de 20 milhões de recursos digitais e o objetivo é que os cidadãos europeus possam aceder a esta informação sem quaisquer restrições. Para não existirem restrições é necessário uma especial atenção às funcionalidades de tradução e à sua eficácia.

No caso da Europeana continua a existir a seleção das línguas disponibilizadas para traduzir todo o interface. Não são traduzidas em nenhuma língua de estudo todos as denominações dos campos de metadados. Uma denominação que permanece sempre em inglês ao aplicar a funcionalidade de tradução é “Providing country”. (ANEXO XII) Ainda existe uma funcionalidade de tradução especificamente direcionada para o recurso. Está funcionalidade destina-se a traduzir o conteúdo textual de descrição dos metadados com o auxílio do “Microsoft Translator”, trata-se de um serviço em linha de tradução por máquina, ou seja, tradução automática. Esta funcionalidade disponibiliza ao todo 44 línguas para tradução. Comparando este número de línguas com o número de línguas no qual se encontra disponível o interface, ou seja, 29, levanta-se de novo a questão se a disponibilização de diferentes línguas, inexistentes numa função mas existentes noutra, é pertinente. Ainda outra observação que se considera necessária fazer nesta linha de raciocínio é que se chega a ter três línguas no mesmo interface: os conteúdos não traduzidos que permanecem em inglês, a língua selecionada, o português e a descrição do objeto em si, que nem com a funcionalidade de tradução adicional são traduzidos, ou seja, o alemão que será a língua de proveniência do objeto digital. Isto é apenas um exemplo mas alastra-se a todo um universo de recursos. Trata-se aqui de uma situação na qual a importância dos elementos não traduzidos, em termos de coerência, é essencial. Estes aspetos podem interferir em todo o processo de pesquisa e limitar o utilizador no processo de recuperação de informação. A Europeana destaca que “ao disponibilizar serviços de acesso multilinguístico, os conteúdos devem estar

acessíveis de forma igualitária para todos os utilizadores da plataforma, e isto independentemente da sua língua nativa e da língua dos recursos” (PETRAS, 2011, p.5-7). No entanto a única língua que é disponibilizada de forma completa é o inglês enquanto as outras línguas apresentam as incoerências já referidas. Por fim, os recursos digitalizados textuais, não se encontram traduzidos, no espaço desta biblioteca digital, nem existe qualquer funcionalidade que permitam a tradução dos objetos. Esta pode ser uma prática que de momento ainda não é possível, mas encontra-se entre os planos futuros da Europeana.

c)WDL

As línguas disponíveis no interface da representação de um recurso específico são idênticas às disponibilizadas nos outros interfaces. Seja qual for a língua aplicada, a tradução do interface é completa. Mesmo os textos informacionais disponibilizados pela biblioteca e uma funcionalidade áudio desses mesmos textos têm acesso igualitário em todas as línguas que esta biblioteca digital disponibiliza. Todos os metadados e elementos textuais encontram-se igualmente traduzidos. O documento digitalizado é que não é alvo de nenhuma funcionalidade de tradução. Não existem quaisquer observações negativas a fazer em relação a esta forma de apresentação dos recursos. O recurso está completamente acessível e o utilizador não tem qualquer problema em consultar todos os aspetos para resolver a sua necessidade de informação. (ANEXO XIII)

5- Funcionalidades Web 2.0

a) ICDL, Europeana e WDL

Não existe qualquer funcionalidade que possa ser considerada como fazendo parte do paradigma da Web 2.0. No caso da ICDL existe no interface principal no tópico “The Foundation” uma ligação designada “For Translators”. Entrando nessa ligação dentro da Informação para tradutores é possível o voluntariado para tradutores que podem candidatar-se para ajudar na tradução dos metadados e do interface (ANEXO XIV). Existe também uma lista com todos os tradutores que já se encontram a colaborar divididos nas várias línguas pelas quais são responsáveis. Isto demonstra não só a não utilização de tradução automática e assim uma certa preocupação com a qualidade dos conteúdos. No entanto pela forma pouco óbvia, espontânea e intuitiva em

que esta funcionalidade se encontra disponibilizada não pode ser considerada nem comparada com uma funcionalidade de *crowdsourcing*.

No entanto existe um reconhecimento geral da necessidade da colaboração com os utilizadores para a tradução e a melhoria dos conteúdos disponibilizados. A própria Europeia destaca que nos seus futuros planos enquadra-se uma plataforma que integrará métodos de *user-assisted query translation* e *crowdsourcing* que apelam à interação ativa dos utilizadores. (EUROPEANA, 2012, p.7) Também a WDL demonstra preocupação neste sentido referindo que valoriza e pretende uma abertura para a participação dos utilizadores. (OUDENAREN, 2009)

3.2. Análise comparativa

Esta análise comparativa tem como foco principal criar uma linha lógica de análise entre os três estudos de caso. O objetivo neste âmbito é concluir esta análise com o levantamento de boas práticas que facilitam o acesso e a recuperação da informação multilingue.

Em termos de funcionalidades multilingues e após a observação e análise de todos os casos, pode afirmar-se que em todos eles se depreendeu a necessidade de várias opções no que diz respeito à disponibilização de línguas, integrar fatores multilinguísticos nos índices de pesquisa e criação de estratégias para deixar o utilizador lidar de forma intuitiva com os recursos. Em todas as bibliotecas existe um grande foco no utilizador e no seu público-alvo. Contudo, a concretização de todas estas preocupações pode apresentar algumas falhas.

No caso da ICDL a adequação ao público-alvo é discutível pois insere um público infantil. Este pormenor é importante destacar pois o ideal seriam funcionalidades de pesquisa e navegação simples de utilizar e intuitivos. Pode então considerar-se que todas as funcionalidades de pesquisa e navegação, com grande variedade, e toda a complexidade conferida pela discrepância de línguas, que não é coerente, um pouco excessiva. O fator da incoerência e a falta de simplicidade pode causar dificuldade no acesso e a recuperação da informação especialmente para um público infantil. A adequação da Europeia ao seu público-alvo, ou seja, principalmente a um público europeu (embora tenha importância internacional) é conseguida pois integra as línguas necessárias. No entanto, a disponibilização mais completa em inglês

revela alguma inconsistência, uma vez que se trata de um público-alvo multilíngue por excelência. No caso da WDL, que tem como alvo um público internacional e multicultural, o único ponto de crítica poderá ser a, ainda, disponibilização de poucas línguas. Porém, qualquer das línguas disponibilizadas está perfeitamente apta e completa para responder às necessidades de informação.

No entanto conclui-se que, em todos os casos, realmente existe consciência da sua missão em fornecer aos utilizadores as ferramentas necessárias para que estes possam proceder às tarefas de acesso e recuperação da informação em todos os casos. Porém, embora as funcionalidades existam, como se verificou, não é garantido que funcionem de forma adequada e, desta forma, é possível que elementos de relevância essencial, intermédia e secundária não sejam traduzidos. Tal ocorre, especialmente no caso do ICDL, onde até foi destacada a não tradução da designação dos metadados que é algo essencial para a compreensão do próprio conteúdo textual dos mesmos. Estes elementos não traduzidos foram considerados de importância essencial e a sua não tradução influencia negativamente o acesso à informação. Na Europeana também existem algumas falhas na tradução, podendo destacar-se o título dos recursos na representação dos resultados, ficando sempre o título original. Já na WDL existe uma preocupação nítida com a tradução de todos os pormenores. A tradução completa de todos os conteúdos, embora uma meta que seja difícil de atingir, devido ao esforço técnico e de mão-de-obra que necessita, é um grande fator qualitativo neste âmbito, porque o que se pretende em todos os casos é uma experiência igualitária para todos os utilizadores independentemente da sua cultura e língua. No caso da Europeana e da ICDL verifica-se ainda que a versão mais forte do sistema é a inglesa, o que dificulta a sua afirmação como bibliotecas digitais multilíngues.

No entanto não são apenas estes fatores que levam ao “quebrar” da fronteira linguística. A disponibilização coerente destas funcionalidades é igualmente importante. Neste campo já se encontram diferenças substanciais entre os estudos de caso. No que diz respeito à ICDL, como já foi destacado, não existe uma linha coerente na experiência da recuperação de informação linguística. O interface de acesso está disponibilizado em menos línguas que os outros interfaces e na pesquisa já é possível o acesso a um grande número de línguas. Porém, entrando na dimensão do recurso, as línguas variam de objeto para objeto. Estas características tornam o ambiente da própria biblioteca digital bastante instável e pode ser referido como um caso de excesso de

personalização em termos de línguas disponíveis. Já na Europeia este problema também se põe mas em mais pequena escala com a disponibilização de duas ferramentas de tradução uma do próprio sítio e ainda a utilização da tradução por máquina, ou seja, automatizada, podendo no fim o utilizador ser confrontado com a representação de um recurso em três línguas diferentes. A WDL parece ser a única que, apesar de não disponibilizar um número tão grande de línguas como a ICDL e a Europeia, segue uma linha coerente de acesso e recuperação da informação. A WDL não apresenta desigualdade linguística no acesso e discrepâncias entre línguas como se verificou nos outros projetos. Neste âmbito a filosofia de menos línguas mas com mais qualidade é perfeitamente adequada. Uma aposta na simplicidade mas eficiência é aquilo que mais é demonstrada neste caso.

Juntando os dois elementos de grande importância descritiva selecionados para este estudo, a flexibilidade e a coerência, pode afirmar-se que a biblioteca digital que mais se destaca em termos de boas práticas é a WDL. Embora no caso da ICDL seja evidente um nítido esforço da disponibilização de muitas funcionalidades de pesquisa e navegação existindo uma grande panóplia de oportunidades de recuperação de informação, estas não se traduzem na prática da melhor forma e muitas vezes até se apresentam inadequadas. Contudo, não pode ser ignorado o seu mérito pelo grande investimento e pelo grande número de funções e possibilidades que disponibiliza. Como referido na revisão de literatura trata-se de um projeto ambicioso em termos de pesquisa e que com mais investimento em termos da melhoria dos conteúdos já existentes poderia apresentar a biblioteca digital multilingue mais completa aqui analisada. No entanto não é isso que se verificou neste estudo apresentando mesmo a necessidade de colmatação de muitas falhas e continuando com a primazia da língua inglesa. A Europeia por estar em constante desenvolvimento ainda está a ser alvo de muitos projetos de melhoria, do qual na perspetiva das funcionalidades multilingue também necessita. Por exemplo, pode referir-se a necessidade da correção das incoerências que se verificarem para dar resposta as necessidades de informação do seu público. No entanto especialmente por surgir no espaço europeu onde a importância da língua é bastante destacada também ela não deveria disponibilizar apenas a versão inglesa na sua forma completa. O WDL é um projeto consistente, coerente, flexível e centrado no utilizador que apresenta todos os benefícios já referidos. A única crítica possivelmente a efetuar seria a possibilidade da interação com os utilizadores, através das

funcionalidades no âmbito da Web 2.0, pois os seus conteúdos acompanhados de elementos textuais escritos por especialistas e a sua versão áudio seriam um meio perfeito e muito interessante para essa mesma interação.

Passando a análise da eficácia, ou seja, se os objetivos destacados e se o desempenho de cada um dos casos se encontram alinhados com a observação que se fez com o estudo aqui apresentado. É importante considerar aqui os outros valores, como a flexibilidade e a coerência, e a análise feita a cada caso, que contribuem para a apreciação do fator denominado “eficácia”. Em termos da Europeana não se verificou o objetivo da disponibilização da tradução dos conteúdos equivalente em qualquer uma das línguas disponibilizadas pela plataforma. Sendo que, neste âmbito ainda se destacam algumas incoerências. No caso da ICDL que afirma que a sua tradução de metadados se encontra completa e em bom funcionamento também não foi verificado. No entanto, o objetivo comum de quebrar as fronteiras linguísticas é central e está a ser trabalhado, apresentando ainda algumas fraquezas. A finalidade do processo, ou seja, o utilizador aceder à informação de forma generalista encontra-se a ser realizado pois apesar das incoerências destacadas é possível a obtenção de resultados. Contudo, com algumas correções, a experiência da recuperação e acesso à informação por parte do público-alvo destes casos de estudo aqui apresentados seria significativamente melhor. A relação entre utilizador e a biblioteca digital é tanto no caso da Europeana como no caso da ICDL limitada pelas incoerências que apresentam.

No que diz respeito a eficácia, é de destacar o caso da WDL, em termos de desempenho, na sua existência como biblioteca digital multilingue. A sua fraqueza na disponibilização de um reduzido número de línguas comparado com os outros casos está entre as prioridades a longo prazo da biblioteca e será assim, previsivelmente, eliminada. Este projeto pretende ser consistente e foi isso que se verificou. Os metadados são apresentados de forma consistente assim como a descrição de cada recurso e do valor adicional acrescentado aos recursos criados pela própria plataforma. A experiência de navegação e de pesquisa é flexível e adequa-se perfeitamente a qualquer utilizador que usa as línguas disponibilizadas. Como ponto principal e para finalizar esta análise, este é o único caso que fornece ao seu público uma experiência de acesso e recuperação da informação igualitária. Isto reflete mais uma vez um dos pontos cruciais incluídos no âmbito da Unesco que referem quatro pilares importantes: a criação do conhecimento, a preservação, partilha e aplicação do mesmo. Desta forma

pretendem evitar fatores de infoexclusão, criar desenvolvimento e potenciar o acesso igualitário à informação. (ABID, 2009, p. 1)

A partir do estudo de cada caso individualmente e através da comparação dos vários casos foi possível denotar que é realmente a WDL que possui as melhores práticas, ao nível do acesso e da recuperação da informação que se assumem como as mais adequados tendo sempre em conta o universo estudado.

As boas práticas passíveis de serem identificadas são:

- Tradução completa: todas as línguas disponibilizadas estão representadas de forma igual, promovendo desta forma o acesso igualitário, não preferencial, de um determinado idioma.
- Flexibilidade, coerência e consistência em termos de usabilidade: facilidade na utilização das ferramentas de tradução oferecidas. Estas são igualmente coerentes pois não existe discrepância nas línguas oferecidas, nem nas diferentes funcionalidades de pesquisa ou nos diferentes interfaces. Esta coerência é o que torna esta biblioteca digital um espaço que pode ser considerado consistente.
- Simplicidade: as utilizações das funcionalidades multilingues são disponibilizadas de forma intuitiva. Existem mecanismos inseridos nos índices de pesquisa e na filtragem dos resultados de fácil utilização.
- Alinhamento com os objetivos do projeto: no caso de uma biblioteca digital multilingue é determinante que os objetivos da biblioteca estejam alinhados tanto nas línguas escolhidas como na forma como a disponibilização multilingue é efetuada.

Estas boas práticas aqui identificadas são consideradas essenciais para uma RI eficaz no âmbito das bibliotecas digitais multilingues. São critérios básicos para oferecer uma experiência de acesso à informação agradável aos utilizadores. No entanto, é também necessário ter em conta que se trata de um resultado adquirido por estes três estudos de caso selecionados, sendo que dos três um se destacou. As boas práticas foram retiradas deste caso, ou seja, da WDL, uma biblioteca digital multilingue que se considerou um excelente exemplo desta tipologia.

4.A rentabilização do multilinguismo em bibliotecas digitais

Os aspetos fulcrais desta dissertação prendem-se, essencialmente, com três pilares denominados avaliação, foco no utilizador e interdisciplinaridade, isto é, a avaliação através do estabelecimento de critérios de análise, a sua utilização e o reconhecimento de boas práticas. Por outro lado, é importante destacar o desenvolvimento de bibliotecas digitais multilingues e das funcionalidades próprias deste campo, bem como a avaliação das mesmas para efeitos de melhoria contínua. Os utilizadores são o foco pela sua importância quando se trata da temática da avaliação. São estes que devem estar no centro de qualquer tipo de avaliação, seja esta de natureza mais técnica ou de usabilidade, assumindo um lugar mais preponderante com o aproveitamento do conhecimento gerado pela sua ação. Este conhecimento pode ser agregado e rentabilizado, traduzindo-se numa melhoria significativa para qualquer sistema de informação. O ponto fulcral no que diz respeito à interdisciplinaridade é o reconhecimento de que, para a construção de um sistema de acesso e recuperação da informação multilingue, é necessário um trabalho interdisciplinar de várias áreas, nomeadamente de natureza linguística e de terminologia.

Bibliotecas digitais multilingues estão a ser criadas por países com mais de uma língua oficial, pela União Europeia e por consórcios de investigação e empresas internacionais. (BISWAS, 2005, P.178) Tal verificou-se, igualmente, na escolha dos estudos de caso. Atualmente a grande maioria dos projetos digitais multilingues tem origem na disponibilização de riqueza e património cultural. A comunidade de informação europeia e internacional, com o seu meio envolvente multilinguístico, apresenta um grande desafio em termos políticos, económicos e de inclusão social. O acesso à informação de modo igualitário é um valor essencial no qual se centra grande parte da investigação nesta área, para combater a realidade da infoexclusão. Especialmente importante para o desenvolvimento futuro é a criação de um mercado digital unificado, especialmente no que diz respeito à sua importância económica e política. (META Technology Council, 2012, p.3) O multilinguismo e os desenvolvimentos adjacentes criam necessidade de acesso igualitário à informação por parte de todos os cidadãos. Como foi referido anteriormente, a exclusão informativa ainda é uma realidade e um grande tema de discussão. A língua mais presente e mais poderosa continua a ser o inglês, que assume sem dúvida, nos dias de hoje, a liderança. De um ponto de vista ideal, qualquer biblioteca digital e qualquer catálogo em linha

deveria disponibilizar uma versão em várias línguas, tendo obviamente em conta a sua localização geográfica, os seus objetivos e a realidade social na qual a biblioteca se insere. Deste modo, poderia alinhar estes fatores com as funcionalidades multilíngues disponibilizadas. Uma biblioteca digital multilíngue, por definição, possibilita uma escolha da língua para os interfaces de acesso, independentemente dos documentos disponibilizados. (PAVANI, 2001, p.73) Tal noção deve ser integrada em todos os projetos deste género sendo que o acesso e a recuperação de informação não deveriam ser limitados pela insuficiência de línguas disponibilizadas. A RI deveria ser independente deste fator, concentrando-se nas necessidades dos utilizadores. Deste ponto de vista é importante destacar a afirmação de BISWAS (2005, p.176) segundo o qual, o ideal será uma biblioteca digital global com capacidade de acessibilidade em todas as línguas. Este tipo de biblioteca obviamente ainda não existe, mas é importante a investigação e o desenvolvimento neste sentido. A investigação na área de MLIA e o desenvolvimento de novos aspetos técnicos para melhorar os aspetos práticos é essencial. A partir de mapeamento semântico, de indexação, da criação de vocabulários controlados, da tradução automática e de dicionários integrados, cada vez mais a informação é tornada acessível. É também de considerar que cada língua apresenta desafios diferentes, sendo que, por exemplo, línguas que não se guiam pelo alfabeto latino até geram dificuldade na apresentação. É importante referir isto pois o acesso global à informação é uma realidade desejável mas também uma realidade ainda utópica. O desenvolvimento de um sistema que suportasse todas as línguas e formatos sem qualquer limitação e falhas não parece um objetivo, neste momento, credível. Estas dificuldades podem observar-se mesmo no estudo de caso aqui efetuado, no que diz respeito ao facto de muitas vezes existirem conteúdos que não são traduzidos, especialmente nos casos da ICDL e da *Europeana*. Se já a tradução completa de todos os conteúdos nas línguas disponibilizadas apresenta um desafio, pois estamos a falar de um meio onde os recursos estão em constante crescimento. Então cada língua e a sua tradução ainda trazem desafios maiores e mais complexos. Torna-se no entanto óbvio que estes elementos vêm interferir com a eficácia da recuperação de informação e com o acesso à mesma. Tal acontece, tanto no caso da ICDL como no caso da *Europeana* - tanto a apresentação dos resultados da pesquisa, como a apresentação de um recurso específico são limitados devido a falhas a nível linguístico e técnico inter-relacionadas. Mas como SORG (2011, p.9) afirma, é improvável, com uma grande coleção de recursos em variadas línguas e de diferentes domínios, garantir a qualidade da RI. Por

consequente, melhorar o desempenho das funcionalidades de pesquisa e navegação torna-se uma tarefa complexa. Já o caso da WDL, ao seguir fatores coerentes e flexíveis, disponibilizando todos os conteúdos de forma completa em qualquer das línguas presentes, não apresenta falhas significativas. Embora estando longe de corresponder ao ideal, é esta a biblioteca digital apresentada que permite ao utilizador ter acesso e recuperar a informação disponibilizada de forma flexível sem limitações significativas nas estruturas apresentadas ao público.

Porém, é também de destacar que o multilinguismo no que diz respeito à área do acesso e da recuperação de informação tem vindo a ganhar relevância. Considera-se que essa relevância, especialmente considerada pelas bibliotecas digitais multilingues, resulta em fatores de inspiração em sistemas como o Google. Estes sistemas permitem um acesso automático, interativo e intuitivo a uma grande variedade de línguas e a tradução de praticamente todos os conteúdos, obviamente sem critérios de qualidade. No entanto é um sistema aceite e com adeptos em todo o mundo, possibilitando uma experiência de RI bastante personalizável. É previsível que as bibliotecas digitais se deixem fascinar pelas funcionalidades que o Google apresenta, aproximando-se deste tipo de sistema, sendo que os projetos digitais que disponibilizam conhecimento de forma controlada têm que se adaptar ao ambiente no qual eles se inserem: a internet.

Para tornar os sistemas que oferecem essas funcionalidades mais “user-friendly”, é necessário a avaliação e o estabelecimento de critérios de análise que permitem, de formas simples, adaptar as funcionalidades multilingues oferecidas ao utilizador. O que se verificou no desenvolvimento da revisão de literatura e no estabelecimento dos critérios de análise foi a falta de critérios de avaliação já existentes, especificamente adaptados à avaliação de funções multilingues no âmbito das bibliotecas digitais. A avaliação destas funcionalidades é essencial, pois tem como objetivo uma melhoria que pode beneficiar toda a biblioteca digital, especialmente fatores de usabilidade. O desenvolvimento de critérios de análise específicos, avaliando minuciosamente todos os aspectos, da visualização à tradução, seria benéfico não só para os criadores de tais plataformas mas, principalmente, para os utilizadores.

O utilizador, em qualquer biblioteca digital e em qualquer método de avaliação, é sempre um elemento determinante de forma direta ou indireta. Para além disso, nesta dissertação foi seguida especificamente uma linha de orientação que confere grande

importância aos utilizadores de bibliotecas digitais multilingues. Nesta perspetiva, para além de se focar a promoção de um acesso igualitário à informação, promove-se também a participação dos utilizadores na criação e melhoria dos recursos disponibilizados. Esta participação por parte dos utilizadores pode ser realizada aproveitando as funcionalidades da Web 2.0. Como já foi destacado, estruturas e funções interativas em ambientes dinâmicos como, por exemplo, o *crowdsourcing* podem contribuir para fins como a melhoria e a criação da informação mas também podem surgir como aspeto de personalização, levando o utilizador a criar uma relação com o espaço digital que consulta. Desta forma seria possível o comentário a traduções já existentes, corrigir as mesmas e criar novas traduções. Também a chamada de atenção para falhas do sistema ou erros de outra natureza podiam ser manifestados por este meio. Nos estudos de caso aqui tratados apesar de serem projetos de importância internacional, não se verificou o aproveitamento dos serviços que a Web 2.0 pode proporcionar. Tal pode dever-se ao facto de a monitorização dessa atividade consumir algum tempo, sendo também que uma biblioteca é por excelência um espaço onde os vocabulários controlados são de extrema importância. É necessário continuar a garantir que estes ambientes de interação adotados pelas bibliotecas possam ser controlados e monitorizados para, desta forma, rentabilizar a interação com os utilizadores e adaptá-la às necessidades da biblioteca digital multilingue. SORG (2011, p.7) afirma que o uso de *Social Semantics* parece indicado para MLIR. Esta adequação é motivada pelas propriedades dos conjuntos de dados que origina. As fontes dos dados recolhidos neste âmbito devem seguir alguns critérios, assim como suportarem todas as línguas e conexões entre estas, realidade que se apresenta na sua maioria no ambiente da Web 2.0. Devem ainda cobrir um grande número de tópicos, pois sistemas RI raramente se encontram limitados por domínios. Em consonância, portais de Web 2.0 normalmente também não se encontram limitados pelo tópico. Outro critério é a atualização pois sistemas de RI necessitam de se adaptar rapidamente a novos tópicos. Consequentemente, os recursos de dados necessitam de ter as mesmas propriedades. No campo de ação da Web 2.0 os conteúdos estão em constante crescimento e são mais adaptáveis a novos tópicos.

Encontrar meios de criação, aproveitamento e rentabilização no âmbito das bibliotecas das funcionalidades Web 2.0, especialmente no que se trata de aspetos de melhoria e criação, é essencial para futuros desenvolvimentos. A aposta na

personalização, na interatividade, na dinamização e na utilização intuitiva dos serviços é essencial para a evolução da tipologia das bibliotecas digitais e para a sobrevivência destas no meio competitivo que é a internet. É também de destacar aquilo que SORG afirma, isto é, a informação recolhida deve ser em todas as línguas e cobrir uma grande variedade de tópicos. Todos estes elementos estão em constante atualização, ou seja, é necessário conseguir lidar com essa informação fornecida e conseguir tratá-la convenientemente para poder integrá-la. Especialmente no caso da tradução de conteúdos e sugestões de melhoria são necessários mecanismos de controlos.

Para garantir aspetos qualitativos é necessário apostar em áreas como a interdisciplinaridade. Neste caso pretende-se dar ênfase à necessidade que existe em reconhecer a ligação entre aspetos terminológicos e a importância que estes têm no multilinguismo em ambientes digitais. A melhor forma de análise e de trabalho no âmbito multilingue surge de uma abordagem interdisciplinar. Esta abordagem interdisciplinar deve sempre englobar tanto terminólogos como especialistas de um determinado domínio, neste caso da ciência da informação. No âmbito da internet trata-se de um ambiente multilingue e multicultural e qualquer abordagem de um projeto de qualquer domínio perseguido neste contexto deve ser de carácter multilingue. (SILVA, 2012, P.76) Só a consciência de que mesmo numa simples tradução é necessário ter em conta vários aspetos, tanto a nível do conceito que se pretende transmitir, como na correta tradução linguística de um determinado termo, trata-se de uma tarefa complexa que exige conhecimentos culturais e linguísticos. A consciência tem ainda de ser recíproca, tanto do lado terminológico, tendo sempre em vista para que tipo de plataforma de informação estão a contribuir, como do lado da ciência da informação, ao ter consciência destas particularidades e como estas podem interferir no acesso e na recuperação da informação. Aqui é necessário alertar para a questão de que a qualidade da tradução disponibilizada nos estudos de casos tratados nesta dissertação não foi passível de análise. Embora seja determinante garantir a qualidade, tanto linguística como conceptual, tendo em conta diferenças culturais em termos de tradução, é igualmente importante que isto seja destacado no âmbito da avaliação. Tal como o estudo de caso referido na revisão de literatura, de avaliação humana de resultados de tradução por máquina presente na revisão de literatura, que reconhece a importância e a dimensão qualitativa de aspetos tradutórios no âmbito de sistemas de acesso e recuperação da informação, a análise de casos de bibliotecas digitais multilingues já

existentes pode ajudar na identificação de boas práticas que podem representar uma base de apoio determinante para projetos já existentes e futuros.

Em resumo existe noção da complexidade que implica a implementação de funcionalidades multilingues no espaço de uma biblioteca digital, mas também se reconhece a urgência que estas funcionalidades representam no acesso e na recuperação da informação. A avaliação destas funcionalidades é uma tarefa que deveria ser reconhecida mesmo em métodos de avaliação global, especialmente no que se refere a fatores como a usabilidade e qualquer critério relacionado com o campo do utilizador. Idealmente, também pontos mais técnicos devem ser avaliados, como algoritmos e formas de conferir cada vez mais rentabilidade em termos de resultados ao sistema. Também aspetos como a tradução e a adequação desta ao público-alvo, aos conteúdos integrados e a funções disponibilizados deve ser objeto de avaliação. Este estudo e os critérios de análise aqui dispostos não são suficientes para avaliar o problema global do multilinguismo em bibliotecas digitais. Contudo, podem ser úteis para a colmatação de falhas que interferem diretamente no processo de acesso à informação por parte do utilizador.

Conclusões

Em conclusão, pode ser afirmado que uma biblioteca digital é um ambiente holístico, no qual qualquer alteração efetuada tem repercussões determinantes noutras áreas da biblioteca. Desta forma, muitas das decisões tomadas seja a nível técnico, prático ou teórico podem influenciar a interação e a relação que o utilizador estabelece com uma biblioteca digital multilingue. A avaliação insere-se no âmbito deste sistema holístico, fazendo parte do mesmo como um fator externo determinante, que confere uma melhoria, ou tem como consequência a melhoria e a alteração de certos fatores fundamentais para uma biblioteca digital. Em especial, a avaliação no âmbito do acesso e da recuperação da informação é vital pois permite, para além da melhoria continua e da correção, o realinhamento dos objetivos definidos para cada projeto.

No espaço de bibliotecas digitais multilingues, as propostas de avaliação no âmbito das funcionalidades de RI, apresentam algumas limitações, nomeadamente pela falta de projetos já existentes e de casos de aplicação. Em primeiro lugar, os métodos para avaliar estas funções, mesmo em estruturas de avaliação mais generalistas, são ainda muito escassos. Em muitos casos, a avaliação limita-se aos resultados em termos de tradução, o que, não obstante ser determinante, revela a falta de avaliações mais completas e a necessidade de investigação nesta área. No mesmo sentido, as possibilidades de exploração da Web 2.0 para o fim de contribuírem para o multilinguismo em plataformas digitais de RI carecem de exploração, verificando-se também aqui a falta de aplicações práticas, não sendo ainda a sua aplicação uma realidade comum e generalizada. A Web 2.0 não se destina apenas a promover os serviços de uma determinada instituição. Ela também pode contribuir para o enriquecimento da mesma. No entanto, também estas funcionalidades, quando existentes, devem ser objeto de avaliação.

Tendo em atenção a importância dada à avaliação, esta dissertação tem como objetivo contribuir para este campo destacando alguns critérios que podem ajudar a identificar e alertar para incoerências no acesso e na recuperação da informação. O multilinguismo é um ponto fulcral nesta área, não só por se inserir no campo da internet que disponibiliza o conhecimento nas mais variadas línguas, mas também para os utilizadores que podem vir de meios culturais e linguísticos diferentes.

A parte essencial desta dissertação são os seus resultados, tanto os critérios de análise criados a partir da observação e da revisão de literatura, como a sua aplicação aos três estudos de caso selecionados. Esta aplicação consistiu numa análise detalhada de cada caso, seguindo-se uma análise comparativa. Dessa comparação foi possível destacar a biblioteca digital multilingue que demonstrou ser a mais eficaz, disponibilizando as suas funcionalidades multilingues de forma coerente e flexível, tendo sempre como foco o utilizador: a *World Digital Library*. Tendo em conta esta situação, foi possível identificar algumas boas práticas que são outro resultado desta dissertação. Entre estas práticas podem destacar-se a disponibilização da tradução completa dos conteúdos, o caráter flexível, coerente e consistente em termos de apresentação multilingue da informação, a utilização intuitiva das ferramentas linguísticas disponibilizadas e o alinhamento dos objetivos do projeto multilinguístico com as opções linguísticas e de disponibilização tomadas. Estas boas práticas foram identificadas, contrastando com as falhas destacadas. Entre estas falhas pode referir-se a discrepância da disponibilização de línguas nas funcionalidades multilingues, falta da tradução de metadados e de outros conteúdos não traduzidos e a continuação do domínio da língua inglesa. Estes fatores têm como consequência, a incoerência, a falta de flexibilidade e, por fim, falhas no que diz respeito a eficácia.

Destacou-se na elaboração da dissertação tanto a importância do multilinguismo, como a importância do utilizador e da avaliação. Foi também destacado o ponto fulcral da interdisciplinaridade, traduzindo-se como a importância de aspetos linguísticos e terminológicos para a área da RI multilingue. Todos estes fatores contribuem para um fim imprescindível: a importância do acesso igualitário à informação.

Referências bibliográficas

ABID, Abdelaziz (2009) – The World Digital Library and Universal Access to Knowledge [Em linha]. [Consult. 13 de Jan. 2013]. Disponível em: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/programme_doc_wdl.pdf

AMATO, Giuseppe, [et al.] (2007) - MultiMatch- Multilingual/Multimedia Acces to Cultural Heritage [Em linha]. [Consult. 27 de Jan. 2013]. Disponível em: <http://www.multimatch.eu/docs/IRCDLSavino.pdf>

ARCHANA, M; DEVI, Sumithra (2011) – Multilingual Information Retrieval Based On Knowledge Creation Techniques. In International Journal of Computer Science, Engineering and Information Technology (IJCSEIT). V.1, nº 4. [Consult. 20 Fev. 2013]. Disponível em: <http://airccse.org/journal/ijcseit/papers/1011ijcseit03.pdf>

BARVE, Sunita (2012) - An evaluation of open source software for building digital libraries [Em linha]. [Consult. 24 Abr. 2013]. Disponível em: <http://shodhganga.inflibnet.ac.in/handle/10603/3731>

BISWAS, Subal Chandra (2005) – Multilingual Access to Information in a networked environment character encoding & Unicode standard. In 3rd Convention PLANNER [Em linha]. p.176- 186. [Consult. 12 Fev. 2013]. Disponível em: <http://ir.inflibnet.ac.in/bitstream/handle/1944/1391/26.pdf?sequence=1>

BUDZISE-WEAVER, Tina; CHEN, Jianping; MITCHELL, Mikhaela (2011) – Collaboration and Crowdsourcing: The cases of multilingual Digital Libraries [Em linha]. [Consult. 18 Jan. 2013]. Disponível em: <http://txcdk-v10.unt.edu/MRT/publications/2012-EL-2.pdf>

CASTAGNÉ, Michel (2012) – Measuring Up: Na Exploration of Digital Library Evaluation Models and Methodologies. [Consult. 27 Maio 2013]. Disponível em: http://web.ncf.ca/castagne/papers/diglib_evaluation.pdf

CHEN, Jiangping; BAO, Yu (2009) – Information Access across Languages on the Web: From search engines to digital libraries [Em linha]. [Consult. 13 Jan. 2013].

Disponível em: <http://max.lis.unt.edu/homepage/publications/2009-Conference-ASIST.pdf>

CHEN, Jianping; [et al.] (2012) – Integrating Machine Translation with Digital Collections for Multilingual Information Access [Em linha]. [Consult. 21 Mar. 2013]. Disponível em: <http://amta2012.amtaweb.org/AMTA2012Files/papers/Chenetal.pdf>

DIEKEMA, Anne (2012) – Multilinguality in the Digital Library: A review. In The electronic Library [Em linha]. V. 2, nº 30, p. 165-181. [Consult. 13 Jan. 2013]. Disponível em: http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1007&context=itls_research

EUROPEANA (2012) – Midterm report on Innovative Multilingual Information Access. European think culture [Em linha]. [Consult. 24 Jan. 2013]. Disponível em: <http://pro.europeana.eu/documents/866067/b0103ac0-611f-4a04-b4b6-f28e200a04e1>

EUROPEANA (2013) – Search [Em linha]. [Consult. 12 Ago. 2013]. Disponível em: <http://www.europeana.eu/>

FRANKLIN, Brinley; KYRILLIDOU, Martha; PLUM, Terry (2009) – From usage to user: library metrics and expectations for the evaluation of digital libraries. *In Evaluation of Digital Libraries: An insight into useful application and methods*. Oxford: Chandos Publishing. ISBN: 97884334 484 1. P. 17-39

GÄDE, Maria (2011) – User Behavior and Evaluation of Multilingual Information Access in Digital Libraries [Em linha]. [Consult. 24 Mar. 2013]. Disponível em: <http://www.ieee-tcdl.org/Bulletin/v7n1/papers/gaede.pdf>

GORNOSTAY, Tatiana; [et al.] (2012) – Consolidating European Multilingual Terminology across Languages and Domains [Em linha]. [Consult. 12 Mar. 2013]. Disponível em: <http://www.ep.liu.se/ecp/072/007/ecp12072007.pdf>

GÖDERT, Winfried (2007) – Multilingualität und Lokalisierung zur Wissenserkundung oder Vom Nutzen semantischer Netze für das Information Retrieval [Em linha]. (2007). [Consult. 24 de Jan. 2013]. Disponível em: http://linux2.fbi.fh-koeln.de/crisscross/publikationen/goedert_bregenz_2006_textfassung.pdf

HALL, Lindsey Elizabeth (2010) – International Children’s Library: A Library for the World’s Children [Em linha]. [Consult. 3 Fev. 2013]. Disponível em: http://coe.winthrop.edu/jonesmg/LTI/2010Fwhitepapers/hall_lindsey.pdf

HUTCHINSON, Hilary Browne; [et al.] (2006)– The International Children’s Digital Library: A Case Study in Designing for a Multilingual, Multicultural, Multigenerational Audience [Em linha]. [Consult. 11 Fev. 2013]. Disponível em: <http://hcil2.cs.umd.edu/trs/2004-24/2004-24.pdf>

ICDL (2010?) – International Children’s Library (ICDL) [Em linha]. [Consult. 3 Fev. 2013]. Disponível em: <http://www.ala.org/aasl/sites/ala.org.aasl/files/content/conferencesandevents/confarchive/pittsburgh/EXPLICDL.pdf>

INTERNATIONAL CHILDREN’S DIGITAL LIBRARY (ICDL) (2013) – International Children’s Digital Library: A Library for the World’s Children. [Consult. 2 Ago. 2013]. Disponível em: <http://en.childrenslibrary.org/>

JENG, Judy (2009) – What should we take into consideration when we talk about usability. In Evaluation of Digital Libraries: An insight into useful application and methods. Oxford: Chandos Publishing. ISBN: 97884334 484 1. p.63-73.

MAEDA, Akira (2002) – Multilingual Information Processing for Digital Libraries [Em linha]. [Consult. 2 Jan. 2013] Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=583E66E1D5576A54AB4B5EB65434FFD5?doi=10.1.1.116.7446&rep=rep1&type=pdf>

MAKARA, Mabofokeng (2010) – Incorporating multilingual bibliographic records into library databases: the case of Lesotho. In World Library and Information Congress: 76th IFLA General Conference and Assembly[Em linha]. [Consult. 23 Mar. 2013]. Disponível em: <http://conference.ifla.org/past/2010/93-makara-en.pdf>

MENDES, Olga Bessa (2008) – A Gestão da Informação na Biblioteca Especializada- As linhas de fronteira da pesquisa de informação. Dissertação de Mestrado em Terminologia e Gestão da Informação de Especialidade. Dissertação do mestrado Terminologia e Gestão da Informação de Especialidade. Lisboa: [s.n.].

META Technology Council (2012) – Strategic research agenda for multilingual Europe 2020 [Em linha]. [Consult. 24 Maio 2013]. Disponível em: http://www.meta-net.eu/vision/reports/meta-net-sra-version_1.0.pdf

MINERVA – Multilingual Access to the European Cultural Heritage: Multilingual Websites and Thesauri. Minerva Plus Project [Em linha]. (2006). [Consult. 7 Nov. 2012]. Disponível em: http://www.minervaeurope.org/structure/workinggroups/inventor/multilingua/documents/Multilingualism_v1_printed.pdf

OUDENAREN, John Van (2009) – The World Digital Library and the Social Science. In World Library and Information Congress: 75th IFLA General Conference and Council[Em linha]. [Consult. 23 Mar. 2013]. Disponível em: <http://conference.ifla.org/past/2009/217-oudenaren-en.pdf>

PAVANI, Ana (2001) – A model of multilingual digital library. In Ciências da Informação. V. 30, nº 3. [Consult. 13 Jan. 2013]. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000300010

PETRAS, Vivien (2011) – Multilingual Information Access in Digital Libraries: Report on Multilingual Access Strategies to Digital Libraries. Europeana connect [Em linha]. [Consult. 27 Out. 2012]. Disponível em: http://www.europeanaconnect.eu/documents/D2.7.1_eConnect-Facilitation%20and%20exchange%20of%20multilingual%20access%20strategies%20to%20digital%20libraries_v1.0.pdf

SARACEVIC, Tefko (2009) – Introduction: the Framework for digital library evaluation. In Evaluation of Digital Libraries: An insight into useful application and methods. Oxford: Chandos Publishing. ISBN: 97884334 484 1. p. 1-15

SAUBERER, Gabriele (2011) – There is no knowledge without terminology, how terminological methods and tools can help to manage monolingual and multilingual knowledge and communication. In Systemics, Cybernetics and Informatics. V. 9, nº 2. [Consult. 26 Mar. 2013]. Disponível em: [http://www.iiisci.org/journal/CV\\$/sci/pdfs/OL787QV.pdf](http://www.iiisci.org/journal/CV$/sci/pdfs/OL787QV.pdf)

SILVA, Manuel Fernando Moreira da Silva (2012) – Localização de Ontologias de Domínio no Contexto de Redes Colaborativas. Dissertação de doutoramento na Especialidade de Lexicologia, Lexicografia e Terminologia. Lisboa: [s.n.].

SORG, Philipp (2011) - Exploiting Social Semantics for multilingual Information Retrieval [Em linha]. [Consult. 20 Mar. 2013]. Disponível em: <http://d-nb.info/1014279925/34>

SUJATHA, Pothula; DHAVACHELVAN, P. (2011) – A Review on the Cross and Multilingual Information Retrieval. International Journal of Web & Semantic Technology [Em linha]. Vol. 2, nº4. [Consult. 24 Jan. 2013]. Disponível em: <http://airccse.org/journal/ijwest/papers/2411ijwest09.pdf>

VULLO, Giuseppina (2010) – A Global Approach to Digital Library Evaluation. In Liber Quaerterly. V. 2, nº 20. [Consult. 27 Maio 2013]. Disponível em: <http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-113589/8301>

WOLDERING, Britta (2006) – Connecting with users: Europe and multilinguality [Em linha]. [Consult. 20 Jan. 2013]. Disponível em: http://www.valaconf.org.au/vala2006/papers2006/95_Woldering_Final.pdf

WORLD DIGITAL LIBRARY (2012) - The World Digital Library: Project Manager's Report [Em linha]. [Consult. 23 Fev. 2013]. Disponível em: http://project.wdl.org/meeting_2012/PM_Report_December2012.pdf

WORLD DIGITAL LIBRARY (WDL) (2013) – Biblioteca Digital Mundial. [Consult. 13 Ago. 2013]. Disponível em: <http://www.wdl.org/pt/>

ZHANG, Ying (2010) – Developing a Holistic Model for Digital Library Evaluation. In Journal of the American Society for Information Science and Technology. V. 1, nº61, p. 88-110. [Consult. 27 Maio 2013]. Disponível em: <https://comminfo.rutgers.edu/~tefko/Courses/Zadar/Doctoral%20studies/Zhang%20dl%20evaluation%20JASIST%20from%20dissertation%202010.pdf~>

ZHOU, Yilu (2006) – Supporting Multilingual Internet Searching and Browsing [Em linha]. [Consult. 3 Mar. 2013]. Disponível em: http://arizona.openrepository.com/arizona/bitstream/10150/195311/1/azu_etd_1763_sip1_m.pdf